

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 08:09:23 ; Search time 3973.1 Seconds

(without alignments)  
19326.012 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689

Sequence: 1 atgaaagtcacgcgtgctt.....tcttcctgttaaaaaaaaaa 5689

Scoring table: OLIGO\_NDC  
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.\*

1: em\_estbda:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlm:\*  
5: em\_estloy:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	683	12.0	767	10	BI917860
2	625	11.0	725	10	BE792557
3	589	10.4	1150	10	BM466811
4	582	10.2	870	10	BG743714
5	563	9.9	713	10	BG170486
6	536	9.4	536	9	A1079538
7	536	9.4	536	9	A1497808
8	536	9.4	536	9	A1038061
9	530	9.3	608	9	AM452651
10	523	9.2	592	9	AL599694
11	522	9.2	574	10	BG470230
12	517	9.1	645	10	BE693704
13	492	8.6	599	9	AM780417
14	487	8.6	809	10	BI093826
15	487	8.6	575	10	BI870563
16	486	8.5	486	9	AI952086
17	484	8.5	486	9	AI333325

18	478	8.4	660	10	BG473458
19	474	8.3	694	10	BG910244
20	463	8.1	463	9	AI356239
21	462	8.1	462	9	AI309934
22	457	8.0	476	9	AA291351
23	447	7.9	966	10	BG260547
24	447	7.9	1063	10	BM462581
25	445	7.8	517	10	BE881579
26	441	7.8	987	10	BG170054
27	436	7.7	436	9	AI192867
28	436	7.7	442	9	AI990048
29	428	7.5	600	10	BE299145
30	427	7.5	427	9	AA293275
31	426	7.5	429	9	AA136967
32	426	7.5	477	10	BG283908
33	422	7.4	452	10	BF508040
34	418	7.3	539	9	AI146646
35	417	7.3	830	10	BI464081
36	409	7.2	1040	10	BG745723
37	403	7.1	446	9	AI369277
38	396	7.0	560	10	BI861416
39	395	6.9	403	9	AI865808
40	395	6.9	545	9	AI129988
41	394	6.9	394	9	AI241357
42	391	6.9	497	9	AA307843
43	388	6.8	569	10	BF061374
44	387	6.8	431	9	AA313349
45	367	6.5	540	9	AM242451

## ALIGNMENTS

RESULT 1  
LOCUS BI917860 767 bp mRNA linear EST 16-OCT-2001  
DEFINITION 603183990P1 NIH\_MGC\_121 Homo sapiens CDNA clone IMAGE:5247718 5',  
mRNA sequence.  
BI917860  
BI917860.1 GI:16199788  
EST.

ACCESSION human.  
VERSION Homo sapiens  
KEYWORDS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
SOURCE NIH-MGC  
ORGANISM NIH-MGC  
REFERENCE NIH-MGC http://mgs.ccl.nhl.gov/  
AUTHORS 1 (bases 1 to 767)  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-rt@mail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM11625 row: d column: 23  
High quality sequence stop: 759.

## FEATURES

source  
1..767  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5247718"  
/clone\_id="NIH\_MGC\_121"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1; NotI;  
Site: 2; EcoRV (destroyed); RNA source anonymous pool of 3  
fetal brains, female age 20 weeks, female age 24 weeks,  
and male age 26 weeks. Library is oligo-dT primed and  
directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.7 kb, insert size range

Page 2

SOURCE  
 ORGANISM  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE  
 AUTHORS  
 TITLE  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 JOURNAL  
 COMMENT  
 Unpublished (1999)  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [gcapsb@mail.nih.gov](mailto:gcapsb@mail.nih.gov)  
 Tissue procurement: DCTD/DPH  
 CDNA Library Preparation: DCTD/DPH  
 DNA Sequencing by: The I.M.A.G.E. Consortium  
 Clone distribution by: Incyte Genomics, Inc.  
 found through the I.M.A.G.E. Consortium/LNLN at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LILCM/88 row: b column: 11  
 High quality sequence stop: 725.  
 Location/Qualifiers  
 1..725  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3939370"  
 /clone\_id="NIH-MGC-7"  
 /tissue\_type="small cell carcinoma"  
 /cell\_line="MGC3"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; CDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GCCACGAC(G) Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 256 a 118 c 207 g 144 t



```
Db 421 ACAAGATATAAATGAAGCCAAAGGAGANTGCTGAAGGGCTTGGAGACATGTTCCAGT 480
Oy 3077 ttgcaaacatcgaagaagatgcaagattgagaaaaacggttaaaataataacaggaat 3136
Db 481 TTGGCAACATCGAAGAAAGATGACAGATTGACAAAACGGGTAAATATAATACAGAAAT 540
Oy 3137 ccttaacatcagaagaagagatacgaaatgaagcaggagacaggaagagattcaagca 3196
Db 541 CTTTACATTCAGAAAGAGAGAGATACGAAATGACAGCAGCAGCAGAGAGATTTCAAGCCA 600
Oy 3197 aaaccggaatttaggaagacagacagctcgagagcgtfaccatctgtgaattcaagatt 3256
Db 601 AAACCTCGAATTTAGGGAACGACAAAGCTCGAGAGCGTACTATGCTGAATTCAGATT 660
Oy 3257 ttcatcgacattgg 3272
Db 661 TTCAATCGAGCATTTGG 676

RESULT 3
LOCUS BM466811 1150 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT 6431318 NIH_MGC_67 Homo sapiens CDNA clone IMAGE:5503072
ACCESSION BM466811
VERSION BM466811.1 GI:18515853
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1150)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE NIH-MGC Consortium of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM12142 row: d column: 17
High quality sequence start: 9
High quality sequence stop: 652.
High quality sequence stop: 652.
Location/Qualifiers
FEATURES
SOURCE 1..1150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5503072"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site: 1; NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 294 a 310 c 309 g 235 t 2 others
ORIGIN
Query Match 10.4%; Score 589; DB 10; Length 1150;
Best Local Similarity 99.8%; Pred. No. 1.2e-268;
Matches 639; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3333 cgttagacctcagagccagaggaagagcatalgtatgcttgtatgcccagatcaa 3392
Db 38 CGCTAGACCTCAGAGCCACGAGGAGCATATGATGATCTTGTATGCTCAAGTCAA 97
Oy 3393 gaagccgaggaattccaaacctcctgttagacagtaacgatacctcagacatca 3452
|||||
```

```
Db 98 GAAGCCGGGAATTCGAAACCCCTACCTGTAGACAGTAAACAGATCACTCTAGCAATCA 157
Oy 3453 tgatcgatacagcgtctcgtgagcaagaatttcagcaagcaagaagatgaatgata 3512
Db 158 TGATCGATACACCGCTCTAGGACAAATTTTCAGCAACCAACAGATGAAATGTAGA 217
Oy 3513 agatcgtcggcgagcactaagattttgaagcaaccttggccgaacgacgcccggcagca 3572
Db 218 AGATCGTCGGCGAGACTATAGTTTGTAGCAACCCCTGGCCGAACGACGCGCGCAGCGCA 277
Oy 3573 gaggcgagcactcgtgttcgtgaggtgcaatgcaagcggcagcaggaagagcg 3632
Db 278 GAGCGGGCCACACTCGGTCTCGTGAAGTGCAAGATGCAAGCGGACGCGCAGAGAGAGG 337
Oy 3633 cgagagctcccgacagcagcagcagcagcagctctcgtcctcggcgaagcaggaanaa 3692
Db 338 CGAGAGCTCCACAGACAGGCCAGCGCAGTACAGCTCTCTCCGCAAAAGCAGAAAAA 397
Oy 3693 tgcagctcgtctcccaagactcttggagcagaactactcccttggggaagcttcca 3752
Db 398 TCCACACTCGGTCTCCAGAGACTTTGGAGAGCAGAACTACTCCCTGGGGAAGCTTTCCA 457
Oy 3753 ggtgccaagaaggaaccccaagctaccagctaccagagctccaggaagagctactgag 3812
Db 458 GAGTGCCTCAAGAGAACCCAGGTAATCTCAGTACCAAGGCTCCAGGAGAGGCTTGGG 517
Oy 3813 aggacatgctccaagcagcagcagctcgtggaactcaggaagctcctcgcaggaaga 3872
Db 518 AGGACATGCTCTCAACGCCAGGCTCATGCTGCAAACTCAGAGCTCTTTCGCCAGAGACA 577
Oy 3873 gaggcggaagagcagcagcagatgaagaagcagctcctcgcaggaagcagcagcagatga 3932
Db 578 GAGGCGGAGAGAGACAGACAGATGAGAAAGCAGCCTCTTCGAGGGGCCACGAACTATGA 637
Oy 3933 ctctgataagaagctcaggaagcagcagcagctcagcagcagcagcagcagcagcagc 3972
Db 638 CTCGTATAGAAAGTCCAGAGCCCGAGTACGCCCTCC 677

RESULT 4
LOCUS BG743714 870 bp mRNA linear EST 15-MAY-2001
DEFINITION 602632895F1 NCI_CGAP_Skn3 Homo sapiens CDNA clone IMAGE:4778302 5',
mRNA sequence.
ACCESSION BG743714
VERSION BG743714.1 GI:14054367
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 870)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10633 row: m column: 23
High quality sequence start: 5
High quality sequence stop: 844.
High quality sequence stop: 844.
Location/Qualifiers
FEATURES
SOURCE 1..870
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4778302"
/clone_lib="NCI_CGAP_Skn3"
```

Query Match  
Best Top

High quality sequence at: [BLAST](#) at: [BLAST](#)

```
/organism="Homo sapiens"
/db_xref="taxon:"
```

```
/clone= "IMAGE:4425975"
```

```

tissue_type="hypernephroma"
lab_host="NIH_MGC_89"

```

Site 2: Saitama, cell line  
DH10B (phage-resistant)  
Organ: kidney; Vector

...: 3.0 kb; Cloned unidirectionally; Average insert size 1.3 kb; null-length

Note: this is a NIH MGC clone and constructed by T. 167

206 g 127 t

5.9%; Score 563; DB 10; Length  
99.9%; Pred. No. 3

```
0; Mismatches 0; Indels
```

CGGCTAATTTTTCAGGAATCCTTACATCAAA

CCGTAATAATAACAGGAATCTTACATCAGAA

AGCAGGCTTCAAGCCAAACTCGAATTT

5'-GAGCAGGAGAGGATTCAGCCAAACTCGAGAAATTT-3'

cyacctaagtgaattcaagatttcacgcgaacatt+

AGACATGCTGAATTCAGATTTCATCGGACATTT

-gggggagttcttcttcttatgaaaggttccatgactat  
 |||||  
 -ggg-

GGGCGAGTTCTTCTTATGACGTTCCATGCCGTCTA-33cclccaa

gaaggcattatgatgatgcttctgtatgccact

5AAGGCATATGATGATGC-TTGATGCGCAATAT

ccacctgtagacagtaacagatcaactctctaa

CACTGTAGACAGTACAGATCAACTCTG

aaagaatttcagcaagcaagcaagatgaatt

AGAAATTTCAGCAAGCAAGCAAGATCTA

...GAAGATCTAGAAG  
tga gca acc cct ggc gac cgc acg ca cgc ggc

TGAGCAACCTGGCCGACGCACGCCCGC

gaggtgcagatgcagcgcgaacccctt

GAGGTGCAGATG CAGCGCAGCCCA

...GAGGAGGAGCGCGAG

AGTACAGCTCTGCGTCGCAAGCAGGAAATGCG

5'-TCCGCAAGCAGGAAATGCC

33GAGCAGACTACTCCGCGA

.....CTGGGGAAGCTTCCAGAGTCTC

DB	3759	caaaagaaacccagctactccag	3782
Db	660	CAAAAGAAACCCAGCTACTCCAG	683
RESULT	6		
LOCUS	AI079538/c		
DEFINITION	AI079538	536 bp	mRNA linear EST 29-SEP-1998
ACCESSION	AI079538		
VERSION	AI079538.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Mumaiyola; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: c9apbs@email.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1447 Std Error: 0.00 Seq primer: -40m3 fwd. ET from Amersham High quality sequence stop: 445. Location/Qualifiers 1..536 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1674370" /clone_lib="Soares_fetal_liver_spleen_INFLS_S1" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pTZ19 (Pharmacia); digested with Pac I and cloned into the Pac I sites of the modified pTZ19 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."		
FEATURES	source		
BASE COUNT	174 a 126 c 114 g 122 t		
ORIGIN			
Query Match	9.4%; Score 536; DB 9; Length 536;		
Best Local Similarity	100.0%; Pred. No. 2.0e-243;		
Matches 536; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
5144	agccgttacacatcacacagagaacaatttcgggtccacagcgggtgtgtggaagaagct	5203	
536	AGCCGTTACACATCACACAGAGAACATTTTCGGGTCCACACACGGGTGTGCGAGAAAGCT	477	
5204	caactcgcgtacagattagagtggtgtgtgtgtcgcggagatcgttggtcccatc	5263	
476	CACCTCTCGCGTACAGATTAGACTGTGTGTGTGTGCGGTCTCGGGATCTCGTGCTCCATC	417	
5264	ttctctcatgttctgaacatctcttatgttaacacatggtcgtgggtgctaaagtgc	5323	
416	TTCTCTCATGTTCTTGAACATCTCTTATGTGAACACATGCGTGGGTGCTAAAGTCCG	357	
5324	tbaattcccgatgtgaaaaaagctggaagtggaagtcagcatcacatgattactttaa	5383	
356	TCBAATCCCGATGTGAAAAAGCTGGAAGTGGAAGTCCACATACCATGATTATCTGTTAA	297	

OY	5384	aaacgaaaaaaaagcagatgtctggtaagtcatctatttttcttatggaacaattgat	5443
Dd	236	AAACGAAAAAAGCAGATGATGGATATCTATTATTTTTTTTATTTGGCACATTGTATTT	237
OY	5444	tctgttgacttgtttttaagaataatgatgtgcacacacagtcaccgtgtctcttcgat	5503
Dd	236	TTCGTGTGACTTGTTTTTAGAAATGATGTGTCACACAGTACC GGTCCTTCGTGCAT	177
OY	5504	tctgtgtcaatggtctcgctttccctaatacagtgcgagggtgtcctaagtgtgttacga	5563
Dd	176	TTCGTGTGACATGGTCTCTTTCTTATATACGGCGCGCGGTCTCTTAAGTGGTACCA	117
OY	5564	gtacgcagatgaccttgatgatacagtgagtccttcttcacagccccctcgaagctgag	5623
Dd	116	GTAACGCCAGATGACCTTTGGATGACAGTGGCTTTTCTACAGCCCTCCCTGAAGCTGT	57
OY	5624	aaacagcttctctgttacatalgcaactctccaataaaaggatatltctctgtt	5679
Dd	56	AAACAGCTTTCTCTGTACATATGCAACCTTAATAAAGCATATTCTTCTCTGTT	1
RESULT	7		
LOCUS	A1497808/c	536 bp	mRNA EST 14-APR-1999
DEFINITION	tm9ef02.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2165307 3'		
ACCESSION	A1497808		mRNA sequence.
VERSION	A1497808.1	GI:4389790	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.		
TITLE	NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index Unpublished (1998)		
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs-femail.nhl.gov Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.		
COMMENT	CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LMW at: www-bio.liml.gov/botrp/image/image.html Insert Length: 779 Std Error: 0.00 Seq primer: -400P from GlDco High quality sequence stop: 468. Location/Qualifiers 1..536 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_IMAGE:2165307" /clone_id="NCI.CGAP.Brn25" /tissue_type="anaplastic oligodendroglioma" /lab_host="DH10B" /note="Organ: brain; Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTATCAATCTGAAGTGGAGCGCGCGCATGATGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT	174 a	126 c	114 g
ORIGIN			122 t

Query Match  
 Best Local Similarity 9.4%; Score 536; DB 9; Length 536;  
 Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

5144 agccgtttacacatcacacagagagacatttcgtgtccacagcggtgtgtgacaggaagt 5203  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 536 AGCGTTTACACATCACACAGAGACATTTTGGGTCCTCCACAGCGGTGTGTGACAGAGCT 477  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5204 cactctcgcgtacagatagatagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 477  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 476 CACTCTCGGTCAGATATAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5263  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5264 ttccttcattgtctgcgaacatcctgtattgttaacacattgcgtgtgtgtgtgtgtgtgt 417  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 416 TTCTCTTCAATGTCTGAAACATCCGTATGTATTAACATGCTGGGTCCTTAAATGTGCTG 5323  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5324 tgaatcccgatgtgtgaaagagctgtgagtgaaagctgacatcacatgtatttactta 5383  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 356 TGAATCCCGATGTGAAAAAGCTGAGTGAAAGCTACACATGATGATTTACTTTAA 297  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5384 aaacagaaaaaagacatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 297  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 296 AAACAGAAAAAAGACATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 5443  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5444 ttgtgtgactgtgttttagaagaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 237  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 236 TTGTGTGACTGTGTTTAGAAATGATGTGTCCACACAGTACCGATGCTGTCTTCTGTGAT 5503  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5504 ttcctgtatcatgt 177  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 176 TTCTGTGATCATGT 5563  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5564 gtaagcgaatgtacattgtatgtacagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 117  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 116 GTACGCCCAATGTACCTTGTGATGTACAGTGTCTTCTCAGACCTCCCTGACCTGTGAG 5623  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5624 aaacagttctctgtacatgtatgtacatcctcctaataaagacatatttctctgtgt 57  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 56 AAACAGCTTCTCTGTACATGTATGTCAACCTCTAATAAAGCATTTTCTTCTGTGT 1

RESULT 8  
 A1038061/c 561 bp mRNA linear EST 28-AUG-1998  
 LOCUS 0x21b05.x1 Soares\_fetal\_liver\_spleen\_INFLS-S1 Homo sapiens cDNA  
 DEFINITION  
 ACCESSION A1038061  
 VERSION A1038061  
 KEYWORDS EST.  
 SOURCE GI:3277255  
 ORGANISM human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 This clone is available royalty-free through LNC; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: 40m31 fwd. Err from Amersham  
 High quality sequence stop: 465.  
 Location/Qualifiers  
 1..561  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1656945"  
 /sex="male" Soares\_fetal\_liver\_spleen\_INFLS-S1

Query Match  
 Best Local Similarity 9.4%; Score 536; DB 9; Length 561;  
 Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

5119 catgaacataaagaagaagaagccgttacatcacacagagacatttcggtgt 5178  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 561 CATGAACATTAAGAGACAAAGAAAGCCGTTACATCTCACACAGACATTTTTCGGT 502  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5179 ccccaagcgggt 502  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 501 CCCACAGCGGT 5238  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5239 ctggggagatctgt 442  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 441 CTGGGGAGATCTGT 5298  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5299 catggtcgt 382  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 381 CATGGCTGGGT 5358  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5359 tgaagatacattatatttactttaaagaagaagaagaacatgtatgtatgtatgtatgt 322  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 321 TGAAGATACATGATGATTTACTTTAAACAGAAAAAGACATGTATGTATGTCTATTT 262  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5419 ttttttattgtgcacattgtatttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 262  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 261 TTTTATTTTATGTGCACATTTGTATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5478  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5479 acaagttaccctgtct 202  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 201 ACAGTACCCCGTGTCTCTGTGATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5538  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5539 ggggtgtcgaatgt 142  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 141 GGGGTGTCTAAGT 5598  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5599 ctcaagcctccctctgt 82  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 81 CTCACAGCTTCCCTCTGACGT

BASE COUNT  
 176 a 130 c 121 g 134 t

Query Match  
 Best Local Similarity 9.4%; Score 536; DB 9; Length 561;  
 Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

5119 catgaacataaagaagaagaagccgttacatcacacagagacatttcggtgt 5178  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 561 CATGAACATTAAGAGACAAAGAAAGCCGTTACATCTCACACAGACATTTTTCGGT 502  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5179 ccccaagcgggt 502  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 501 CCCACAGCGGT 5238  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5239 ctggggagatctgt 442  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 441 CTGGGGAGATCTGT 5298  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5299 catggtcgt 382  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 381 CATGGCTGGGT 5358  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5359 tgaagatacattatatttactttaaagaagaagaagaacatgtatgtatgtatgtatgt 322  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 321 TGAAGATACATGATGATTTACTTTAAACAGAAAAAGACATGTATGTATGTCTATTT 262  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5419 ttttttattgtgcacattgtatttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 262  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 261 TTTTATTTTATGTGCACATTTGTATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5478  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5479 acaagttaccctgtct 202  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 201 ACAGTACCCCGTGTCTCTGTGATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5538  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5539 ggggtgtcgaatgt 142  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 141 GGGGTGTCTAAGT 5598  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 5599 ctcaagcctccctctgt 82  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 81 CTCACAGCTTCCCTCTGACGT

RESULT 9  
 AWA52651 608 bp mRNA linear EST 17-FEB-2000  
 LOCUS UT-H-B13-bl-u-g-02-0-UI-S1 NC1-CGAP\_Sud5 Homo sapiens cDNA clone  
 DEFINITION  
 ACCESSION AWA52651  
 VERSION AWA52651  
 KEYWORDS EST.  
 SOURCE GI:6993427  
 ORGANISM human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL  
COMMENT

Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaaps-femail.nih.gov  
The sequence contained an oligo-dr track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA library preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/bdrrp/image/image.html](http://www.bio.llnl.gov/bdrrp/image/image.html)  
Seq primer: M13 forward  
POLA-yes.

FEATURES  
source

Location/Qualifiers  
1. 608  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3068738"  
/lab\_host="NCI\_CGAP\_Sub5"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; NCI\_CGAP\_Sub5 is a subtracted library derived from NCI\_CGAP\_Sub4. The NCI\_CGAP\_Sub5 library had 3 million recombinants. A single-stranded DNA preparation of NCI\_CGAP\_Sub4 was used as a tracer in a subtractive hybridization with a driver comprising: the IMAGE pool (NCI\_CGAP\_K1d3 pool 1 LLAM 3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids 1332376-1323911, 1456008-1456775, 1500552-1502855); NCI\_CGAP\_K1d5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids 1323912-1325831, 1471368-1472903, 1492104-1493255); NCI\_CGAP\_Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991, 1520904-1522439); NCI\_CGAP\_GC4 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631, 1469654-1470983, 1475592-1476743); NCI\_CGAP\_P12 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959, 1217928-1220615); NCI\_CGAP\_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255, 114584-1145351). (10% of the driver population), plus a pool of 3,840 arrayed clones from NCI\_CGAP\_Sub1 (IMAGE Clonoids 2708616-2710535) and NCI\_CGAP\_Sub2 (IMAGE Clonoids 2710536-2712455) (10% of the driver population), plus a pool of 11,136 clones from NCI\_CGAP\_Sub3 (IMAGE Clonoids 2712456-2723591) (10% of the driver population), plus a pool of 5,472 clones from NCI\_CGAP\_Sub4 (IMAGE Clonoids 2723592-2728969) (70% of the driver population). Subtraction was performed as previously described [Bonaldi, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6, 791-806.  
TAG\_LIB=NCI\_CGAP\_Brn26  
TAG\_TISSUE=brain  
TAG\_SEQ=ATAGC

BASE COUNT 131 a 154 c 115 g 208 t

Query Match 9.3% Score 530; DB 9; Length 608;  
Best Local Similarity 99.8%; Pred. No. 1.7e-240;

Matches 580; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2411 ccgactgctcttgaagcagatgtgacagcttctgtcttccaagagaagattg 2470  
|||||  
DB 594 CCGACTGCTCTTTAGTCCATGTGATGATGTTCTTTCACAGAGAAGATTG 535  
QY 2471 gaagtcagatgctcagaaacagacaagcaatttcagatgcagtcgaattgat 2530  
|||||  
DB 534 GACGTCAAGATGTCTCAAAAAACGACAAAGCAATTTTCAGATGCCAGTCAATTGATT 475  
QY 2531 tcgttaaaacagaaatcaaaagcagatgttgaagfatagctggaagactaaacca 2590  
|||||  
DB 474 TCGTTAAACACGAAATCAAAAAAGCATGATTAGTATAGCTGACGAGACTAAACTCA 415

QY 2591 atacagtgaatgaccagaagaagcaggttctccacagagaagatgtggctccctgggtc 2650  
|||||  
DB 414 ATACAGTGAATGACCAGAAAGCAGGTCTCCACAGAGAGATGTGGGTCTTCCTGGGGTC 355  
QY 2651 tgaagaatcgaagctcgctgtgagaagcttcgcagaccgagttgcagagtgttgaatg 2710  
DB 354 TGAAAGATTCAGACGCTGTGGAGAGCTTCGACACCGCAGGTGCCAGAGGACTTTGAAG 295  
QY 2711 gggatattccttcgatctgcacagccgcggatataatcaagaaggaggatccaatgaga 2770  
DB 294 GGGATATTCCTTCCATGCTCCACGCCGCGGATATATCAGAGCGAGGAGATCAATGAGA 235  
QY 2771 gcttcagagctgcacatcagcaaatcttatgataaacccgcggtatagatgataaga 2830  
DB 234 GCTTCAGAGCTGCCATTCGACAAATCTTATGATTAACCCGTGTATATGATGATGAAG 175  
QY 2831 gcatgagaaccttggaagaagacacagaagaagaatcagaatcaggaggagagctgtat 2890  
DB 174 GCATGAGACCTTGGAGAGACACAGAAAGTTCAGATCAGGAGAGAGCTGTAT 115  
QY 2891 ccagagcagatgataagcttccactctctgagagaacaaatgaatgaaacaaagaga 2950  
DB 114 CCACGCCAGTATCAGCTTCCTCCACTCTCTGGAGAGACAAATGAGAAACCAAGAGA 55  
QY 2951 aagtgataagactgatagaagaagaagataaactgaaaa 2991  
DB 54 AAGTGATTAAGCTATGAAAAAAGATTAATAACTGAAAA 14

RESULT 10  
AL59694  
LOCUS  
DEFINITION DKF2p313f1730.r1 313 (synonym: h1cc2) Homo sapiens cDNA clone  
KEYWORDS EST.  
ACCESSION AL59694  
VERSION AL59694.1 GI:15162982  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 592)  
AUTHORS Koehner,K., Beyer,A., Mewes,W., Well,B. and Wiemann,S.  
TITLE EST (Koehner,K., Beyer,A., Mewes,H.W., Well,B. and Wiemann,S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Koehner K

MIPS  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
Sequenced by BMFZ (Biomedical Research Center at the Charite,  
Berlin/Germany) within the cDNA sequencing consortium of the German  
Genome Project.

No sl sequence available.  
This clone (DKF2p313f1730) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: [clone@rzpd.de](mailto:clone@rzpd.de).

FEATURES  
source

Location/Qualifiers  
1. 592  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKF2p313f1730"  
/clone\_lib="313 (synonym: h1cc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac; Site\_1: SfiI; Site\_2: SfiI; cDNA-collection"  
BASE COUNT 187 a 141 c 133 g 128 t 3 others  
ORIGIN



REFERENCE  
AUTHORS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 645)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
TITLE  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
COMMENT  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the PAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=et2-NC0-BT0387-100  
800-013-f05&t3=2000-08-10&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 45  
High quality sequence stop: 644.  
Location/Qualifiers  
1..645  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BT0387"  
/dev\_stage="Adult"  
/note="Organ: breast; Vector: puc18; Site:1; Sma1; Site:2;  
Sma1: A mini-library was made by cloning products derived  
from ONESTES PCR (U.S. Letters Patent application No. 196  
, 716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."  
BASE COUNT 122 a 176 c 115 g 232 t  
ORIGIN  
Query Match 9.1%; Score 517; DB 10; Length 645;  
Best Local Similarity 99.8%; Pred. No. 2.6e-234;  
Matches 567; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2510 cagatgcagtcgaattgatttcgtttaaaacgaaatcaaaagcatgattagta 2569  
|||||  
DB 612 CAGATGCCAGTCAATTGCGTTTGTAAACAGAAATCAAAACATGATTTAGTA 553  
QY 2570 tagctgacgagactaaactcaatgacatgacgacgaagaagcatgttccacagag 2629  
|||||  
DB 552 TAGCTGACGAGACTAAACTCAATGACGTGATGACCAAGACAGGTTCTCCAGCAGAG 493  
QY 2630 atgtggctctccctcgtggtctgaaagaagtcgaagctcgttggagagctgcagacgcag 2689  
|||||  
DB 492 ATGTGGGCTCTTCCCGGGCTGTGAAGAACTCAAGCTCATTGGAGAGTCTGCAGACCGCAG 433  
QY 2690 ttgcgcgaggtgacttgaaatgggatatctcttcacatgcgtccacggcgcgagataatca 2749  
|||||  
DB 432 TTGCGCAGGCTGACTTGAATGGGGATATTCTTTTCATCGTCCACGCCGCGGATTAATCA 373  
QY 2750 gaggcagggagatgacatgagatcctcagagctgcacatgacaatcttataaacccg 2809  
|||||  
DB 372 GAGGCGAGGGAATGACATGAGAGCTTCAGAGCTGCATCCACAATTTTATGATAACCCG 313  
QY 2810 cggtagatgatgatgaaggcatggaagccttggaaagaacacagagaagaagtcca 2869  
|||||  
DB 312 CGTAGATGATGATGATGAAGCATGAGACCTTGAAGAAGACACAGAAAGTTCAA 253

QY 2870 gatcagggagagatgtctgataccacagccaagtgatcaagcttccactcttgagagac 2929  
|||||  
DB 252 GATCAGGAGAGAGTCTGTATCCACAGCAGTGCATGATCAGCTTCCACTCTCTGAGAGAC 193  
QY 2930 aatgatgaaacccaaggaaagtgatgaagctatgaagaaaggttaaacggaa 2989  
|||||  
DB 192 AATGATGGAACCAAGAGAAAGTGATGACTATGAAAAAAGATAAAGTGA 133  
QY 2990 aagaagaagaagaatagatagaagaagaagataaaatgaagccaagaaggaatgc 3049  
|||||  
DB 132 AAGAAAAGAAAGATAGATAGATAGAGAGAGATTAATGAAAGCAAGACGATATG 73  
QY 3050 tgaaggcttggagacatgttcaggtc 3077  
|||||  
DB 72 TGAAGGCTTGGCAGACATCTTCAGGTT 45  
RESULT 13  
AW780417/c 599 bp mRNA linear EST 12-MAY-2000  
LOCUS  
DEFINITION  
ho21b08.x1 NCI-CGAP Col4 Homo sapiens cDNA clone IMAGE:3038007 3'  
similar to TR:092340 Q9Z340 ATYPICAL PKC SPECIFIC BINDING PROTEIN.  
; mRNA sequence.  
ACCESSION  
VERSION AW780417 GI:7795020  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 599)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cga@db-rcemail.nih.gov  
Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Possible reversed clone: polyT not found  
Seq primer: -40up from Gibco  
High quality sequence stop: 412.  
Location/Qualifiers  
1..599  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:3038007"  
/clone\_lib="NCI-CGAP-Col4"  
/tissue\_type="moderately-differentiated adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: PCMV-SPOrt6; Site:1; SalI;  
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.7 kb. Life Technologies catalog #:  
11531-019"  
BASE COUNT 130 a 157 c 117 g 195 t  
ORIGIN  
Query Match 9.1%; Score 516; DB 9; Length 599;  
Best Local Similarity 99.8%; Pred. No. 7.9e-234;  
Matches 566; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2422 ttgagtcagatgtgatccagttcttcttcaacgaaagatttgagctcagagt 2481  
|||||  
DB 567 TTGAGTCCAGATGTGATTCAGTTCGTTTCAACGAGAAAGATTGGACCTCAGAGT 508  
QY 2482 atgtcagaagaacgcacaaagaatttccagatgcagtcgaatttgatttcgttaaaaa 2541

Query Match  
 Best Local Similarity 8.6%; Score 492; DB 10; Length 809;  
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 BASE COUNT 252 a 174 c 191 g 192 t  
 ORIGIN

```

Db 507 ATATCAGAAAAACGACAAAGCAATTTTCATGATCCAGTCAATGATTTCTTAACA 448
QY 2542 cgaataatcaaaagatgatttagtataagctacagagactaactaactaactgagat 448
Db 447 CGAAATATCAAAAAGCATGATTTAGTATGATGATGATGATGATGATGATGATGAT 2601
QY 2602 gacacgaagaagcagttctccacagagatgtgtgtgtgtgtgtgtgtgtgtgtgt 388
Db 387 GACCGAAGAAAGAGAGTCTCCAGAGAGATGAGTGTGTGTGTGTGTGTGTGTGTGTGT 2661
QY 2662 agctcgtttgagagctgtgacagcagctgtgacagcagctgtgacagcagctgtgac 328
Db 327 AGCTCATTTGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2721
QY 2722 ttccatgttcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 268
Db 267 TTCCATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2781
QY 2782 ggcacacgaagaatcttatgataaaccgagcagcagcagcagcagcagcagcagcag 208
Db 207 GCCATGCAAAATCTTATGATTAACCCCGGTAGATGATGATGATGATGATGATGATGAT 2841
QY 2842 ttggaagaagacagagagagagagagagagagagagagagagagagagagagagag 148
Db 147 TTGGAAGAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2901
QY 2902 gacacgcttcacacacacacacacacacacacacacacacacacacacacacacac 88
Db 87 GATGACCTTCTCCACTCTCTGTGAGAGACAAATGAAATGAAATGAAATGAAATGAAATG 2961
QY 2962 actgaatgaagaaagagaaactga 2988
Db 27 ACTGATAGAAAAAGATAAACTGGA 1

RESULT 14
BI093826 809 bp mRNA linear EST 20-JUN-2001
LOCUS 602860606F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:5001744 5,
ACCESSION BI093826
VERSION BI093826.1 GI:14512156
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 809)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1035 row: 1 column: 01
High quality sequence stop: 759.
Location/Qualifiers
1. 809
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5001744"
/clone_1ib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

```

```

Db 536 TAGGATTAATGGC 547
QY 2208 taggaataagg 2219
Db 536 TAGGATTAATGGC 547

RESULT 15
BI870563 575 bp mRNA linear EST 11-OCT-2001
LOCUS 603394044F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:5404067 5,
ACCESSION BI870563
VERSION BI870563.1 GI:16044236
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 575)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium/LLNL
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12030 row: 9 column: 12

```



High quality sequence stop: 575.

## FEATURES

Location/Qualifiers

1..575

Source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5404067"  
/clone\_lib="NIH\_MGC\_90"  
/tissue\_type="adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.7 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

BASE COUNT 182 a 139 c 128 g 126 t

ORIGIN

Query Match 8.6%; Score 487; DB 10; Length 575;

Best Local Similarity 100.0%; Pred. No. 5e-220;

Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 925 ggtgtaagcctgaacatgaacaaatcttctcgtgagaatgattgcattgctcagatatt 984
    |||||||
Db 12 ggtgtaagcctgaacatgaacaaatcttctcgtgagaatgattgcattgctcagatatt 71
    |||||||

OY 985 gatggcagccttcgaatagaagatttgaacaagacacaacatatttctcccaagccatg 1044
    |||||||
Db 72 gatggcagccttcgaatagaagatttgaacaagacacaacatatttctcccaagccatg 131
    |||||||

OY 1045 cgtacaaccatcattggttccatggttccctgcagcaaatagaagcagttatgaacaa 1104
    |||||||
Db 132 cgtacaaccatcattggttccatggttccctgcagcaaatagaagcagttatgaacaa 191
    |||||||

OY 1105 ctatcccaagtgagaagaacattactatcaagccgttttagccctgacagccagttat 1164
    |||||||
Db 192 ctatcccaagtgagaagaacattactatcaagccgttttagccctgacagccagttat 251
    |||||||

OY 1165 attgacaacagagtggtgaacagtgcaagggcttcacacggtgcagagagacacccgactg 1224
    |||||||
Db 252 attgacaacagagtggtgaacagtgcaagggcttcacacggtgcagagagacacccgactg 311
    |||||||

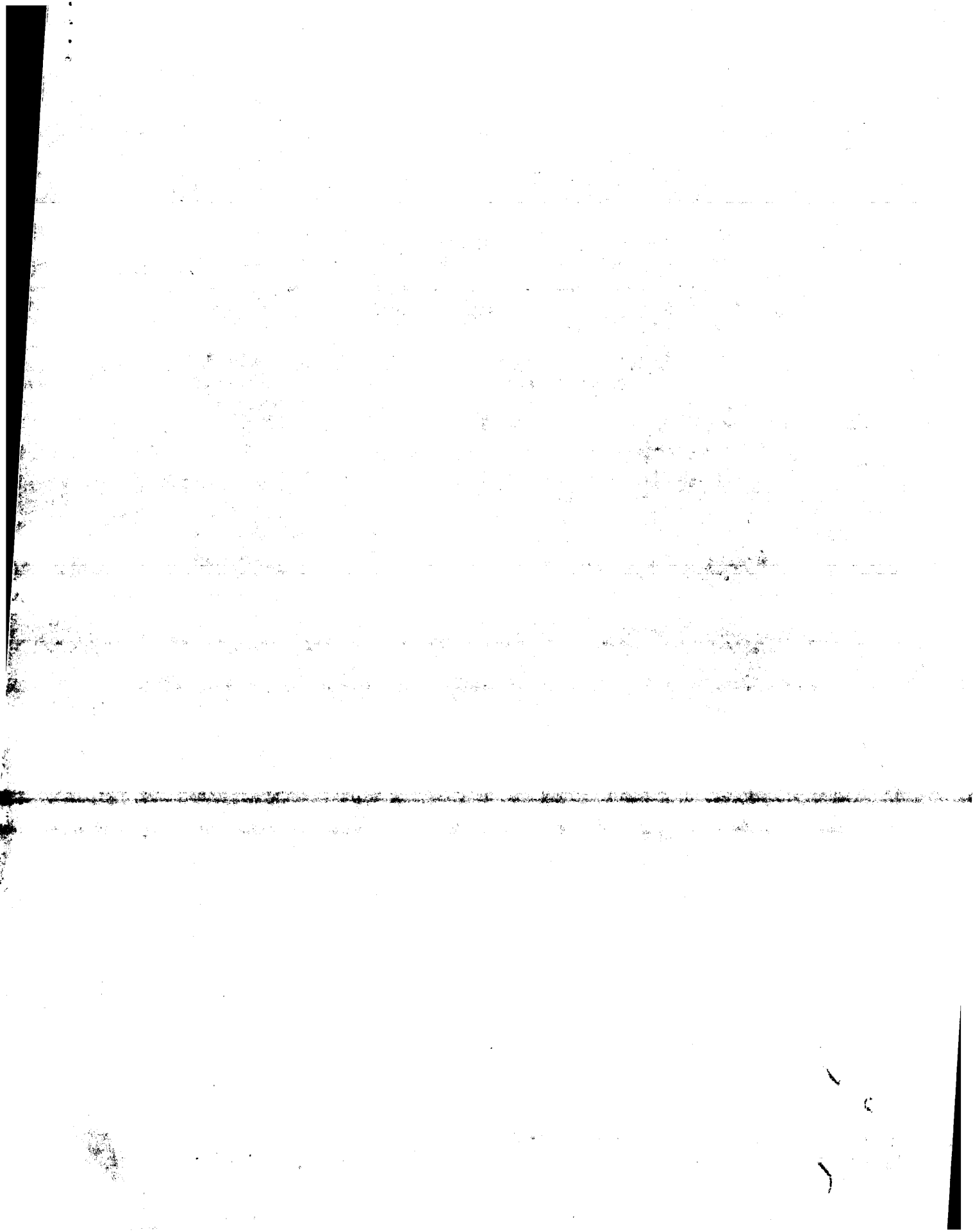
OY 1225 aaccacccgctgagcagatagactctcaactcaagaactcaactcaatagcgacacccctcg 1284
    |||||||
Db 312 aaccacccgctgagcagatagactctcaactcaagaactcaactcaatagcgacacccctcg 371
    |||||||

OY 1285 ggaataacacatcgcgtccagcctcgacactcagaatgtatattagtagactgtaagc 1344
    |||||||
Db 372 ggaataacacatcgcgtccagcctcgacactcagaatgtatattagtagactgtaagc 431
    |||||||

OY 1345 agtggtataacacccaataataagcaagaggtctaataatcagcttaagaaggtaca 1404
    |||||||
Db 432 agtggtataacacccaataataagcaagaggtctaataatcagcttaagaaggtaca 491
    |||||||

OY 1405 gaaggtt 1411
    |||||||
Db 492 gaagcgtt 498
```

Search completed: July 24, 2002, 12:25:52  
Job time: 15389 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 11:12:43 ; Search time 102.97 Seconds

(without alignments)  
13571.030 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689

Sequence: 1 atgaagtgaccgtgtgctt.....tcttcctgttaaaaaaaaaa 5689

Scoring table: OLIGO\_NUC

Searched: 38353 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/1na/PCUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/1na/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	0.4	1272	3	US-08-946-026-36
2	22	0.4	1330	3	US-08-888-077A-26
3	21	0.4	1774	4	US-09-215-252-16
4	19	0.3	39	3	US-08-961-083-371
5	19	0.3	196	3	US-08-961-083-145
6	19	0.3	1027	4	US-09-303-524A-1
7	19	0.3	1628	4	US-08-883-515-3
8	19	0.3	5836	4	US-09-233-086-2
9	18	0.3	565	1	US-08-469-427A-4
10	18	0.3	565	2	US-08-609-443B-4
11	18	0.3	561	2	US-08-569-063C-4
12	18	0.3	591	2	US-08-469-427A-6
13	18	0.3	591	2	US-08-609-443B-6
14	18	0.3	591	2	US-08-569-063C-6
15	18	0.3	624	2	US-08-609-443B-12
16	18	0.3	624	2	US-08-569-063C-12
17	18	0.3	886	1	US-08-469-427A-1
18	18	0.3	886	2	US-08-609-443B-1
19	18	0.3	886	2	US-08-569-063C-1
20	18	0.3	966	2	US-08-766-738-2
21	18	0.3	1229	2	US-08-975-316-85
22	18	0.3	1293	3	US-08-722-184-7
23	18	0.3	1293	4	US-09-043-937A-11
24	18	0.3	1343	1	US-08-197-792-30
25	18	0.3	1343	1	US-08-459-850-30
26	18	0.3	1343	1	US-08-459-214-30
27	18	0.3	2563	2	US-08-553-436A-7

c	28	18	0.3	2900	2	US-09-027-337-9	Sequence 9, Appl1
c	29	18	0.3	3153	4	US-09-080-625-3	Sequence 3, Appl1
c	30	18	0.3	3247	4	US-08-718-388-4	Sequence 4, Appl1
c	31	18	0.3	3336	4	US-09-080-625-2	Sequence 2, Appl1
c	32	18	0.3	3694	4	US-09-080-625-5	Sequence 5, Appl1
c	33	18	0.3	3877	4	US-09-080-625-4	Sequence 4, Appl1
c	34	18	0.3	7824	4	US-08-718-388-6	Sequence 6, Appl1
c	35	18	0.3	16382	4	US-08-718-388-8	Sequence 8, Appl1
c	36	17	0.3	300	4	US-09-439-313-240	Sequence 240, Appl1
c	37	17	0.3	523	4	US-08-896-164-13	Sequence 13, Appl1
c	38	17	0.3	570	1	US-08-469-427A-10	Sequence 10, Appl1
c	39	17	0.3	570	2	US-08-609-443B-10	Sequence 10, Appl1
c	40	17	0.3	570	2	US-08-569-063C-10	Sequence 10, Appl1
c	41	17	0.3	573	2	US-08-290-665A-141	Sequence 141, App
c	42	17	0.3	573	5	PCT-US95-10398-141	Sequence 141, App
c	43	17	0.3	771	4	US-09-230-637-8	Sequence 8, Appl1
c	44	17	0.3	950	4	US-08-960-780-12	Sequence 12, Appl1
c	45	17	0.3	950	4	US-09-073-898-12	Sequence 12, Appl1

## ALIGNMENTS

RESULT 1  
US-08-946-026-36  
; Sequence 36, Application US/08946026  
; Patent No. 6034218  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Mitcham, Jennifer L.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND IMMUNODIAGNOSIS OF PROSTATE CANCER  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/946,026  
; FILING DATE: 07-OCT-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MAKI, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.424C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1272 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-946-026-36

Query Match 0.4%; Score 22; DB 3; Length 1272;  
Best Local Similarity 100.0%; Pred. No. 0.81;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3609 gcagcgagcgagcgagcgagcgag 3630  
|||||

us-09-757-781-20.olg.rni

RESULT 2  
US-08-999-077

Sequence 26, Application US/08888077A  
Patent No. 6020143  
GENERAL INFORMATION:  
ADDRESS:

APPLICANT: ROMMENS, HENRIKUS, PETER H  
 APPLICANT: ROMMENS, JOHANNA M  
 TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
 TO ALZHEIMER'S DISEASE AND USES THEREOF  
 NUMBER OF SEQUENCES: 41  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: LERNER, DAVID, LITTENBERG, KRONHOLTZ & MENTLIK  
STREET: 600 SOUTH AVENUE WEST  
CITY: WESTFIELD  
STATE: NJ  
COUNTRY: UNITED STATES

COUNTRY: USA  
ZIP: 07090-1497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION Data:

APPLICATION NUMBER: US/08/888,077A  
FILING DATE: 03-JUL-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/592,541  
FILING DATE: 26-JAN-1996  
ATTORNEY:

NAME: PALISI, THOMAS M  
REGISTRATION NUMBER: 36,629  
REFERENCE/DOCKET NUMBER: SCHERING 3.0-017 CIP CIP CIP IV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 455-1111

TELEFAX: (908) 654-7866  
INFORMATION FOR SEQ ID NO: 26  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1330 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:

NAME/KEY: CDS  
LOCATION: 145..1275  
OTHER INFORMATION: /product="S5a"  
8-888-077A-26

	try Match	0.48;	Score 22;	DB 3;	Length 1330;		
	ct Local Similarity	100.0%;	Pred. No.	0.61;			
	Conservative	0;	Mismatches	0;	Indels	0;	Gaps 0;
3609	gcacgagcgagcggcgaaggagag	3630					
822	GCAGCGCAGCGCGACGAGGAGG	843					

T 3  
 215-252-16  
 Sequence 16 Application US/09215252  
 Invent No. 6300487  
 PUBLICATION INFORMATION:  
 APPLICANT: LEUNG, David W.  
 APPLICANT: ADUREL, Daniel  
 APPLICANT: HOLLENBACK, David  
 DATE OF INVENTION: MAMMALIAN LYSOSEPHATIDIC ACID ACYLTRANSFERASE  
 REFERENCE: 077319/0151  
 FIRST APPLICATION NUMBER: US/09/215,252

```

? CURRENT FILING DATE: 1998-12-18
? PRIOR APPLICATION NUMBER: US 08/618,655
? PRIOR FILING DATE: 1996-03-19
? NUMBER OF SEQ ID NOS: 42
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 15
? LENGTH: 1774
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (158)..(1291)
US-09-215-252-16

```

Query Match	0.4%;	Score 21;	DB 4;	Length 1774
Best Local Similarity	100.0%;	Pred. No. 2.5;		
Matches	21;	Conservative		

	QY	DB	Mismatches	Indels	Gaps
	547	aacactgctggagtcctctaa	0:	0:	0
	871	aacactgctggagtcctctaa	0:	0:	0
		891			

RESULT 4  
JS-08-961-1

Sequence 371, Application US/08961083  
Patent No. 6159469  
GENERAL TNEO

TITLE INFORMATION: APPLICANT: Choi et. al.  
 TITLE OF INVENTION: Streptococcus pneumoniae antigens and Vaccines  
 NUMBER OF SEQUENCES: 452  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA

ZIP: 20850  
COUNTRY: US  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage  
COMPIER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (701) 333-3333

TELEFAX: (301) 309-8504  
INFORMATION FOR SEQ ID NO: 371:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
3-961-083-371

Seq Match	0.34	Score 19	DB 3	Length 39
Local Similarity	100.0%	Prod. No. 23		
Seqs 19	Conservative 0	Mismatches 0	Indels 0	Gaps 0
1159	caaaaataagccaagag	1377		

Db 21 CAAAATAATAGCAGAGG 39

## RESULT 5

US-08-961-083-145  
; Sequence 145, Application US/08961083

; Patent No. 6159469

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,083

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 145:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 196 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; US-08-961-083-145

Query Match 0.3%; Score 19; DB 3; Length 196;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1359 caaaaataagccaagag 1377

DB 11 CAAAATAATAGCAGAGG 29

## RESULT 6

US-09-303-524A-1

; Sequence 1, Application US/09303524A

; Patent No. 6238873

; GENERAL INFORMATION:

; APPLICANT: CHAMBERS, JONATHAN K.

; APPLICANT: STEWART, BRIAN R.

; APPLICANT: AMES, ROBERT S.

; APPLICANT: SARAU, HENRY M.

; APPLICANT: FOLEY, JIM

; APPLICANT: ARNOLD, ANNE ROMANIC

; TITLE OF INVENTION: METHODS OF SCREENING FOR AGONISTS AND

; TITLE OF INVENTION: ANTAGONISTS OF THE INTERACTION BETWEEN THE HUMAN KIAA0001

; FILE REFERENCE: GP50007

; CURRENT APPLICATION NUMBER: US/09/303,524A

; CURRENT FILING DATE: 1999-04-30

; PRIOR APPLICATION NUMBER: 60/083,957

; PRIOR FILING DATE: 1998-05-01

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 1027

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-303-524A-1

Query Match 0.3%; Score 19; DB 4; Length 1027;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 actctgtcagtagtagta 475

DB 420 actctgtcagtagtagta 438

## RESULT 7

US-08-883-515-3/C

; Sequence 3, Application US/08883515

; Patent No. 5981836

; GENERAL INFORMATION:

; APPLICANT: Osteryoung, Katherine W

; TITLE OF INVENTION: PLANT CHLOROPLAST DIVISION GENES

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Quarles & Brady

; STREET: 1 South Pinckney Street

; CITY: Madison

; STATE: WI

; COUNTRY: US

; ZIP: 53701-2113

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/883,515

; FILING DATE:

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Seay, Nicholas J.

; REGISTRATION NUMBER: 27,386

; REFERENCE/DOCKET NUMBER: 920905, 90016

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 608-251-9166

; TELEFAX: 608-251-9166

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1628 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 3..1316

; US-08-883-515-3

Query Match 0.3%; Score 19; DB 2; Length 1628;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 269 gcaccagtcgcaattccac 287

DB 544 GCACCAAGTCGCAATTCAC 526

## RESULT 8

US-09-233-086-2  
Sequence 2, Application US/09233086  
Patent No. 6337192  
GENERAL INFORMATION:  
APPLICANT: Bartel, Paul L.  
APPLICANT: Tavtigian, Sean V.  
TITLE OF INVENTION: Myriad Genetics, Inc.  
FILE REFERENCE: MMS1 - An MMAC1 Interacting Protein  
CURRENT APPLICATION NUMBER: US/09/233,086  
EARLIER FILING DATE: 1999-01-19  
PRIOR APPLICATION NUMBER: US 60/071,861  
NUMBER OF SEQ ID NOS: 65  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 5836  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (115)..(5757)  
US-09-233-086-2

Query Match 0.3%; Score 19; DB 4; Length 5836;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1406 aaggttgagtcagcat 1424  
DB 2195 aaggttgagtcagcat 2213  
|||||

## RESULT 9

US-08-469-427A-4/c  
Sequence 4, Application US/08469427A  
Patent No. 5607918  
GENERAL INFORMATION:  
APPLICANT: Eriksson, Ulf  
APPLICANT: Olofsson, Birgitta  
APPLICANT: Alltalo, Karl  
APPLICANT: Pajusola, Katri  
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND  
TITLE OF INVENTION: DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKewen, Edwards & Lenahan  
STREET: 1200 G Street, N.W., Suite 700  
CITY: Washington  
STATE: DC  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,427A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/397,651  
FILING DATE: 01-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Evans, Joseph D  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 41979cp2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 4:

## SEQUENCE CHARACTERISTICS:

LENGTH: 565 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
TISSUE TYPE: adult mouse heart  
US-08-469-427A-4

Query Match 0.3%; Score 18; DB 1; Length 565;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3608 tgcagcgagcgagcgag 3625  
DB 482 tgcagcgagcgagcgag 465  
|||||

## RESULT 10

US-08-609-443B-4/c  
Sequence 4, Application US/08609443B  
Patent No. 5840693  
GENERAL INFORMATION:  
APPLICANT: ERIKSSON, Ulf  
APPLICANT: OLOFSSON, Birgitta  
APPLICANT: ALLTALO, Karl  
APPLICANT: PAJUSOLA, Katri  
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND  
TITLE OF INVENTION: DNA CODING THEREFOR  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKewen, Edwards & Lenahan, P.L.L.C.  
STREET: 1200 G Street, N.W., Suite 700  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/609,443B  
FILING DATE: 01-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/397,651  
FILING DATE: 01-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/469,427  
FILING DATE: 06-JUN-1995  
APPLICATION DATA:  
APPLICATION NUMBER: US 08/569,063  
FILING DATE: 06-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 1064/41979CP4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 565 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO



```

;
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W., Suite 700
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,443B
; FILING DATE: 01-MAR-1996
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/397,651
; FILING DATE: 01-MAR-1995
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/469,427
; FILING DATE: 06-JUN-1995
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/569,063
; FILING DATE: 06-DEC-1995
;
; ATTORNEY/AGENT INFORMATION:
; NAME: EVANS, Joseph D
; REGISTRATION NUMBER: 26,269
; REFERENCE/DOCKET NUMBER: 1064/41979CP4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
;
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 591 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: adult mouse heart
;
; US-08-609-443B-6

```

```

Query Match          0.3%; Score 18; DB 2; Length 591;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3608 tgcagcgagcgagcgagcg 3625
Db 508 TGCAGCGGCGAGCGGCGAG 491

RESULT 14
; US-08-569-063C-6/c
; Sequence 6, Application US/08569063C
; Patent No. 5928939
; GENERAL INFORMATION:
; APPLICANT: ERIKSSON, Ulf
; APPLICANT: OLOFSSON, Birgitta
; APPLICANT: ALITALO, Kari
; APPLICANT: PATUSOLA, Kari
; TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
; TITLE OF INVENTION: DNA CODING THEREFOR
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W., Suite 700
; CITY: Washington
; STATE: DC
; COUNTRY: USA

```

```

;
; ZIP: 20005
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,063C
; FILING DATE: 06-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/469,427
; FILING DATE: 06-JUN-1995
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/397,651
; FILING DATE: 01-MAR-1995
;
; ATTORNEY/AGENT INFORMATION:
; NAME: EVANS, Joseph D
; REGISTRATION NUMBER: 26,269
; REFERENCE/DOCKET NUMBER: 1064/41979CP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
;
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 591 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: adult mouse heart
;
; US-08-569-063C-6

```

```

Query Match          0.3%; Score 18; DB 2; Length 591;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3608 tgcagcgagcgagcgagcg 3625
Db 508 TGCAGCGGCGAGCGGCGAG 491

```

```

RESULT 15
; US-08-609-443B-12/c
; Sequence 12, Application US/08609443B
; Patent No. 5840693
; GENERAL INFORMATION:
; APPLICANT: ERIKSSON, Ulf
; APPLICANT: OLOFSSON, Birgitta
; APPLICANT: ALITALO, Kari
; APPLICANT: PATUSOLA, Kari
; TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
; TITLE OF INVENTION: DNA CODING THEREFOR
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W., Suite 700
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,443B
; FILING DATE: 01-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```



	Query Match	0.3%	Score 18:	DB 2:	Length 624:
	Best Local Similarity	100.0%	Pred. No. 72:		
Matches	18;	Conservative	0;	Mismatches	0; Indels
Gy	3608	tgcagcgacgacggcagg	3625		
Dd	585	TGCACGGCAGCGCCAGG	568		

Search completed: July 24, 2002, 14:19:22  
Job time: 11199 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 11:13:58 ; Search time 547.74 Seconds  
(without alignments)  
17832.414 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689

Sequence: 1 atgaaagtacccgtgtcctt.....tcttcctgttaaaaaaaaa

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_032802:\*

1: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2991	52.6	5510	AAH73337
2	1694	29.8	2718	AAH18587
3	1694	29.8	2718	AAH78369
4	1004	17.6	1800	AAH72793
5	842	14.8	867	AAH91875
6	725	12.7	823	AAH27320
7	725	12.7	823	AAH33594
8	407	7.2	556	AAH07109
9	395	6.9	425	AAH67583

10	269	4.7	597	22	AAH71480	Human cervical can
11	252	4.4	400	22	AAH66057	Novel human polynu
12	244	4.3	582	23	AAH68901	DNA encoding novel
13	234	4.1	540	21	AAH79294	Human lung tumour-
14	234	4.1	540	23	AAH23370	Human lung tumour-
15	211	3.7	510	23	AAH68902	DNA encoding novel
16	193	3.4	564	22	AAH12111	Human cDNA clone (
17	181	3.2	181	22	AAH74458	Human foetal liver
18	181	3.2	181	22	AAH54924	Probe #23610 used
19	181	3.2	556	22	AAH61963	Human foetal liver
20	181	3.2	556	22	AAH14889	Probe #10575 used
21	181	3.2	1262	23	AAH68903	DNA encoding novel
22	166	2.9	505	22	AAH61506	Human foetal liver
23	166	2.9	505	22	AAH29229	Probe #7695 for ge
24	166	2.9	505	22	AAH09806	Human brain expres
25	166	2.9	505	22	AAH35700	Human bone marrow
26	166	2.9	505	22	AAH17106	Probe #7039 for ge
27	166	2.9	505	22	AAH14115	Probe #10101 used
28	134	2.4	305	22	AAH72113	Human cervical can
29	134	2.4	305	22	AAH71048	Human cervical can
30	120	2.1	504	22	AAH70055	Human cervical can
31	112	2.0	145	21	AAH45600	Human secreted exp
32	110	1.9	339	22	AAH69243	Human cervical can
33	71	1.2	180	21	AAH02894	Human secreted pro
34	52	0.9	2554	21	AAH01182	PD2 domain-compris
35	33	0.6	525	23	AAH67317	DNA encoding novel
36	33	0.6	565	23	AAH68814	DNA encoding novel
37	26	0.5	1090	20	AAH84303	Wheat geminivirus
38	23	0.4	508	22	AAH06145	Human cDNA clone (
39	23	0.4	700	22	AAH00331	Human reproductive
40	23	0.4	700	22	AAH62479	Human breast or ov
41	23	0.4	820	22	AAH14319	Human cDNA sequenc
42	23	0.4	875	22	AAH68551	Human protein HP10
43	23	0.4	127197	22	AAH61370	Soybean 515002 reg
44	22	0.4	396	23	AAH53141	Enterococcus faeca
45	22	0.4	1116	18	AAH78308	E6ap-binding prote

sequence 5510 BP; 1526 A; 1242 C; 1418 G; 1305 T; 19 other;

	Mismatches	0;	Indels	1;	Gaps	1;
--	------------	----	--------	----	------	----

2205 atagcctgacgactaaactcatacagtgatgacccaagaagcaggttctcccaqcaqa 2358

Db 2359 gatgtggtcctccctggtctgaagaaatcaactctttgaacactat

2419 attaccatatttcttccatcgtccacggccgcgataatc 2748

2745 agagggcagggtgcaatgagagcttcaggagctgcatcgacaaatcttatgtataaaccc 2808

Y 2809 gcggtagatgatgatgaaggcatqqaaccttgcgaacacccccc

2869 accatgagcgcacgaaggccctggagaagacacagaaagtcca 2598

2599 agatcaggagagagtcgtatccacagccagtgtatcagcctccgactcttggagaa 3650

2659 caaatqaatgaaacccaaaggaataaaaaggataaaactgya 2988

2205 aaagaaaaagaaagataagagagaagataaatagaagccaagaaggaataa 3048

3049 ctgaaggcttggagacatgttcagatttgcacaaactgctg  
 2778 gaaagccaaagaaaggggaatg

3109 aaaaaccatcaaa  
3110  
3111  
3112  
3113  
3114  
3115  
3116  
3117  
3118  
3119  
3120  
3121  
3122  
3123  
3124  
3125  
3126  
3127  
3128  
3129  
3130  
3131  
3132  
3133  
3134  
3135  
3136  
3137  
3138  
3139  
3140  
3141  
3142  
3143  
3144  
3145  
3146  
3147  
3148  
3149  
3150  
3151  
3152  
3153  
3154  
3155  
3156  
3157  
3158  
3159  
3160  
3161  
3162  
3163  
3164  
3165  
3166  
3167  
3168  
3169  
3170  
3171  
3172  
3173  
3174  
3175  
3176  
3177  
3178  
3179  
3180  
3181  
3182  
3183  
3184  
3185  
3186  
3187  
3188  
3189  
3190  
3191  
3192  
3193  
3194  
3195  
3196  
3197  
3198  
3199  
3200  
3201  
3202  
3203  
3204  
3205  
3206  
3207  
3208  
3209  
3210  
3211  
3212  
3213  
3214  
3215  
3216  
3217  
3218  
3219  
3220  
3221  
3222  
3223  
3224  
3225  
3226  
3227  
3228  
3229  
3230  
3231  
3232  
3233  
3234  
3235  
3236  
3237  
3238  
3239  
3240  
3241  
3242  
3243  
3244  
3245  
3246  
3247  
3248  
3249  
3250  
3251  
3252  
3253  
3254  
3255  
3256  
3257  
3258  
3259  
3260  
3261  
3262  
3263  
3264  
3265  
3266  
3267  
3268  
3269  
3270  
3271  
3272  
3273  
3274  
3275  
3276  
3277  
3278  
3279  
3280  
3281  
3282  
3283  
3284  
3285  
3286  
3287  
3288  
3289  
3290  
3291  
3292  
3293  
3294  
3295  
3296  
3297  
3298  
3299  
3300  
3301  
3302  
3303  
3304  
3305  
3306  
3307  
3308  
3309  
3310  
3311  
3312  
3313  
3314  
3315  
3316  
3317  
3318  
3319  
3320  
3321  
3322  
3323  
3324  
3325  
3326  
3327  
3328  
3329  
3330  
3331  
3332  
3333  
3334  
3335  
3336  
3337  
3338  
3339  
3340  
3341  
3342  
3343  
3344  
3345  
3346  
3347  
3348  
3349  
3350  
3351  
3352  
3353  
3354  
3355  
3356  
3357  
3358  
3359  
3360  
3361  
3362  
3363  
3364  
3365  
3366  
3367  
3368  
3369  
3370  
3371  
3372  
3373  
3374  
3375  
3376  
3377  
3378  
3379  
3380  
3381  
3382  
3383  
3384  
3385  
3386  
3387  
3388  
3389  
3390  
3391  
3392  
3393  
3394  
3395  
3396  
3397  
3398  
3399  
3400  
3401  
3402  
3403  
3404  
3405  
3406  
3407  
3408  
3409  
3410  
3411  
3412  
3413  
3414  
3415  
3416  
3417  
3418  
3419  
3420  
3421  
3422  
3423  
3424  
3425  
3426  
3427  
3428  
3429  
3430  
3431  
3432  
3433  
3434  
3435  
3436  
3437  
3438  
3439  
3440  
3441  
3442  
3443  
3444  
3445  
3446  
3447  
3448  
3449  
3450  
3451  
3452  
3453  
3454  
3455  
3456  
3457  
3458  
3459  
3460  
3461  
3462  
3463  
3464  
3465  
3466  
3467  
3468  
3469  
3470  
3471  
3472  
3473  
3474  
3475  
3476  
3477  
3478  
3479  
3480  
3481  
3482  
3483  
3484  
3485  
3486  
3487  
3488  
3489  
3490  
3491  
3492  
3493  
3494  
3495  
3496  
3497  
3498  
3499  
3500  
3501  
3502  
3503  
3504  
3505  
3506  
3507  
3508  
3509  
3510  
3511  
3512  
3513  
3514  
3515  
3516  
3517  
3518  
3519  
3520  
3521  
3522  
3523  
3524  
3525  
3526  
3527  
3528  
3529  
3530  
3531  
3532  
3533  
3534  
3535  
3536  
3537  
3538  
3539  
3540  
3541  
3542  
3543  
3544  
3545  
3546  
3547  
3548  
3549  
3550  
3551  
3552  
3553  
3554  
3555  
3556  
3557  
3558  
3559  
3560  
3561  
3562  
3563  
3564  
3565  
3566  
3567  
3568  
3569  
3570  
3571  
3572  
3573  
3574  
3575  
3576  
3577  
3578  
3579  
3580  
3581  
3582  
3583  
3584  
3585  
3586  
3587  
3588  
3589  
3590  
3591  
3592  
3593  
3594  
3595  
3596  
3597  
3598  
3599  
3600  
3601  
3602  
3603  
3604  
3605  
3606  
3607  
3608  
3609  
3610  
3611  
3612  
3613  
3614  
3615  
3616  
3617  
3618  
3619  
3620  
3621  
3622  
3623  
3624  
3625  
3626  
3627  
3628  
3629  
3630  
3631  
3632  
3633  
3634  
3635  
3636  
3637  
3638  
3639  
3640  
3641  
3642  
3643  
3644  
3645  
3646  
3647  
3648  
3649  
3650  
3651  
3652  
3653  
3654  
3655  
3656  
3657  
3658  
3659  
3660  
3661  
3662  
3663  
3664  
3665  
3666  
3667  
3668  
3669  
3670  
3671  
3672  
3673  
3674  
3675  
3676  
3677  
3678  
3679  
3680  
3681  
3682  
3683  
3684  
3685  
3686  
3687  
3688  
3689  
3690  
3691  
3692  
3693  
3694  
3695  
3696  
3697  
3698  
3699  
3700  
3701  
3702  
3703  
3704  
3705  
3706  
3707  
3708  
3709  
3710  
3711  
3712  
3713  
3714  
3715  
3716  
3717  
3718  
3719  
3720  
3721  
3722  
3723  
3724  
3725  
3726  
3727  
3728  
3729  
3730  
3731  
3732  
3733  
3734  
3735  
3736  
3737  
3738  
3739  
3740  
3741  
3742  
3743  
3744  
3745  
3746  
3747  
3748  
3749  
3750  
3751  
3752  
3753  
3754  
3755  
3756  
3757  
3758  
3759  
3760  
3761  
3762  
3763  
3764  
3765  
3766  
3767  
3768  
3769  
3770  
3771  
3772  
3773  
3774  
3775  
3776  
3777  
3778  
3779  
3780  
3781  
3782  
3783  
3784  
3785  
3786  
3787  
3788  
3789  
37

[illegible]

2899 aagcaqqaqcaqqaqqaattcaagca 3228

gagcgtgactatgctgaattcaagatttcatcgacattgctgtgatgataa 3288

3289 atgtatggggagtttcttcttataqaaattccatgcctatc  
3018

3078  
 3079  
 3080  
 3081  
 3082  
 3083  
 3084  
 3085  
 3086  
 3087  
 3088  
 3089  
 3090  
 3091  
 3092  
 3093  
 3094  
 3095  
 3096  
 3097  
 3098  
 3099  
 3100  
 3101  
 3102  
 3103  
 3104  
 3105  
 3106  
 3107  
 3108  
 3109  
 3110  
 3111  
 3112  
 3113  
 3114  
 3115  
 3116  
 3117  
 3118  
 3119  
 3120  
 3121  
 3122  
 3123  
 3124  
 3125  
 3126  
 3127  
 3128  
 3129  
 3130  
 3131  
 3132  
 3133  
 3134  
 3135  
 3136  
 3137  
 3138  
 3139  
 3140  
 3141  
 3142  
 3143  
 3144  
 3145  
 3146  
 3147  
 3148  
 3149  
 3150  
 3151  
 3152  
 3153  
 3154  
 3155  
 3156  
 3157  
 3158  
 3159  
 3160  
 3161  
 3162  
 3163  
 3164  
 3165  
 3166  
 3167  
 3168  
 3169  
 3170  
 3171  
 3172  
 3173  
 3174  
 3175  
 3176  
 3177  
 3178  
 3179  
 3180  
 3181  
 3182  
 3183  
 3184  
 3185  
 3186  
 3187  
 3188  
 3189  
 3190  
 3191  
 3192  
 3193  
 3194  
 3195  
 3196  
 3197  
 3198  
 3199  
 3200  
 3201  
 3202  
 3203  
 3204  
 3205  
 3206  
 3207  
 3208  
 3209  
 3210  
 3211  
 3212  
 3213  
 3214  
 3215  
 3216  
 3217  
 3218  
 3219  
 3220  
 3221  
 3222  
 3223  
 3224  
 3225  
 3226  
 3227  
 3228  
 3229  
 3230  
 3231  
 3232  
 3233  
 3234  
 3235  
 3236  
 3237  
 3238  
 3239  
 3240  
 3241  
 3242  
 3243  
 3244  
 3245  
 3246  
 3247  
 3248  
 3249  
 3250  
 3251  
 3252  
 3253  
 3254  
 3255  
 3256  
 3257  
 3258  
 3259  
 3260  
 3261  
 3262  
 3263  
 3264  
 3265  
 3266  
 3267  
 3268  
 3269  
 3270  
 3271  
 3272  
 3273  
 3274  
 3275  
 3276  
 3277  
 3278  
 3279  
 3280  
 3281  
 3282  
 3283  
 3284  
 3285  
 3286  
 3287  
 3288  
 3289  
 3290  
 3291  
 3292  
 3293  
 3294  
 3295  
 3296  
 3297  
 3298  
 3299  
 3300  
 3301  
 3302  
 3303  
 3304  
 3305  
 3306  
 3307  
 3308  
 3309  
 3310  
 3311  
 3312  
 3313  
 3314  
 3315  
 3316  
 3317  
 3318  
 3319  
 3320  
 3321  
 3322  
 3323  
 3324  
 3325  
 3326  
 3327  
 3328  
 3329  
 3330  
 3331  
 3332  
 3333  
 3334  
 3335  
 3336  
 3337  
 3338  
 3339  
 3340  
 3341  
 3342  
 3343  
 3344  
 3345  
 3346  
 3347  
 3348  
 3349  
 3350  
 3351  
 3352  
 3353  
 3354  
 3355  
 3356  
 3357  
 3358  
 3359  
 3360  
 3361  
 3362  
 3363  
 3364  
 3365  
 3366  
 3367  
 3368  
 3369  
 3370  
 3371  
 3372  
 3373  
 3374  
 3375  
 3376  
 3377  
 3378  
 3379  
 3380  
 3381  
 3382  
 3383  
 3384  
 3385  
 3386  
 3387  
 3388  
 3389  
 3390  
 3391  
 3392  
 3393  
 3394  
 3395  
 3396  
 3397  
 3398  
 3399  
 3400  
 3401  
 3402  
 3403  
 3404  
 3405  
 3406  
 3407  
 3408  
 3409  
 3410  
 3411  
 3412  
 3413  
 3414  
 3415  
 3416  
 3417  
 3418  
 3419  
 3420  
 3421  
 3422  
 3423  
 3424  
 3425  
 3426  
 3427  
 3428  
 3429  
 3430  
 3431  
 3432  
 3433  
 3434  
 3435  
 3436  
 3437  
 3438  
 3439  
 3440  
 3441  
 3442  
 3443  
 3444  
 3445  
 3446  
 3447  
 3448  
 3449  
 3450  
 3451  
 3452  
 3453  
 3454  
 3455  
 3456  
 3457  
 3458  
 3459  
 3460  
 3461  
 3462  
 3463  
 3464  
 3465  
 3466  
 3467  
 3468  
 3469  
 3470  
 3471  
 3472  
 3473  
 3474  
 3475  
 3476  
 3477  
 3478  
 3479  
 3480  
 3481  
 3482  
 3483  
 3484  
 3485  
 3486  
 3487  
 3488  
 3489  
 3490  
 3491  
 3492  
 3493  
 3494  
 3495  
 3496  
 3497  
 3498  
 3499  
 3500  
 3501  
 3502  
 3503  
 3504  
 3505  
 3506  
 3507  
 3508  
 3509  
 3510  
 3511  
 3512  
 3513  
 3514  
 3515  
 3516  
 3517  
 3518  
 3519  
 3520  
 3521  
 3522  
 3523  
 3524  
 3525  
 3526  
 3527  
 3528  
 3529  
 3530  
 3531  
 3532

QY	3349	ccacgagaaaggccatagatgagatgctttgtatgcccgaagtcagaagaccggaattcc	3408
Db	3079	ccacgagaaaggccatagatgagatgctttgtatgcccgaagtcagaagaccggaattcc	3138
QY	3409	aaaccctcaaccctgtatgacaagttaaacaagatcctacgacatcatatgctgataagcgt	3468
Db	3139	aaaccctcaaccctgtatgacaagtaacaagaatcaaccctcctagacatcatatgctgataagcgt	3198
QY	3469	ctgagagcagaagattctcagacgaagcaaaagcaagatgaaagatgtatgaagatcgtctccggaacc	3528
Db	3199	ctgagagcagaagattctcagacgaagcaaaagcaagatgaaagatgtatgaagatcgtctccggaacc	3258
QY	3529	tatagttttgggcaaacctctggccggaaccgacgagccgagcgaacgaagggcgagacatcg	3588
Db	3259	tatagttttgggcaaacctctggccggaaccgacgagccgagcgaacgaagggcgagacatcg	3318
QY	3589	gtctccgctggaggtgacagaatgacagccgagccgagcgagggcgagggcgagagctccccaag	3648
Db	3319	gtctccgctggaggtgacagaatgacagccgagccgagcgagggcgagggcgagagctccccaag	3378
QY	3649	gccacagccgacgagacagctctctgtgcctcgcgcaaaagcagaanaatgccaagctgctcc	3708
Db	3379	gccacagccgacgagacagctctctgtgcctcgcgcaaaagcagaanaatgccaagctgctcc	3438
QY	3709	ccagagctcttgggagacgaagatactacccctctgggaaagagcttccagaagtgccaaagaaac	3768
Db	3439	ccagagctcttgggagacgaagatactacccctctgggaaagagcttccagaagtgccaaagaaac	3498
QY	3769	cccacagctacccagctacccaagagctccagaagaaagcgtacctctgggaaagaaatgctctaac	3828
Db	3499	cccacagctacccagctacccaagagctccagaagaaagcgtacctctgggaaagaaatgctctaac	3558
QY	3829	ggccaggtgcatagcctcggaaactcaggaagctctcctccgcgcaagaaacgaagccggaagagcag	3888
Db	3559	ggccaggtgcatagcctcggaaactcaggaagctctcctccgcgcaagaaacgaagccggaagagcag	3618
QY	3889	cagatgaaagaagcagcctcctctccgcggggggcccaagcaactatgactcgtataaagaagctc	3948
Db	3619	cagatgaaagaagcagcctcctctccgcggggggcccaagcaactatgactcgtataaagaagctc	3678
QY	3949	caggagcccccggtaacgcccctcccaaggggccccttcgcgcaaggttggccccctcccct	4008
Db	3679	caggagcccccggtaacgcccctcccaaggggccccttcgcgcaaggttggccccctcccct	3738
QY	4009	tctcaaggttgcgagctcgtacaagacttcaagactcctcgtgaaagggagccctctatcc	4068
Db	3739	tctcaaggttgcgagctcgtacaagacttcaagactcctcgtgaaagggagccctctatcc	3798
QY	4069	tgaagccagcaaatbaacgagctgcttcaatgctgcgcgaataaagaacatttccatitgaagac	4128
Db	3799	tgaagccagcaaatbaacgagctgcttcaatgctgcgcgaataaagaacatttccatitgaagac	3858
QY	4129	tgtatatttggagagtttttcaaaaccctgcatggtactatgaggtatctcgttgttgt	4188
Db	3859	tgtatatttggagagtttttcaaaaccctgcatggtactatgaggtatctcgttgttgt	3918
QY	4189	atccaggtgccccttaagccggtgtaggcgaagaatggaagagcccttaagtctgtttgaagc	4248
Db	3919	atccaggtgccccttaagccggtgtaggcgaagaatggaagagcccttaagtctgtttgaagc	3978
QY	4249	gtctcaaggtgccttcttcatgagaaagatttcccaaccctgtgacaatactcgttttgaagc	4308
Db	3979	gtctcaaggtgccttcttcatgagaaagatttcccaaccctgtgacaatactcgttttgaagc	4038
QY	4309	gttcaatagctctgcgcgctctccacaagctacccaagatccggaccctactaataatcctgaagtggag	4368
Db	4039	gttcaatagctctgcgcgctctccacaagctacccaagatccggaccctactaataatcctgaagtggag	4098
QY	4369	cgcgcgcttgggtttaaaccctcctgcacaacattgcagacttgcaggtgaacaattcactcgaagtggag	4428
Db	4099	cgcgcgcttgggtttaaaccctcctgcacaacattgcagacttgcaggtgaacaattcactcgaagtggag	

Oy 4429 atgacgtgcggcctgttctccctaaagtctcttaagctatgagatgacattagtc 4488  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4159 atgacgtgcggcctgttctccctaaagtctcttaagctatgagatgacattagtc 4218  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 4489 tctgttgagaaagatgaggagacatacacaataatggggcttctcgttacttcaag 4548  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4219 tctgttgagaaagatgaggagacatacacaataatggggcttctcgttacttcaag 4278  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 4549 caccgcatctgtctactctgtcatcactgtgttctctcttccctccagctctt 4608  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4279 caccgcatctgtctactctgtcatcactgtgttctctcttccctccagctctt 4338  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 4609 gtgacggagagatcggtcatccttaataagaagatgacatgcttctgtgt 4668  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4339 gtgacggagagatcggtcatccttaataagaagatgacatgcttctgtgt 4398  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 4669 caccatgggggtctcttctgttaacgctctatgactcaacatataaagtgtatc 4728  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4399 caccatgggggtctcttctgttaacgctctatgactcaacatataaagtgtatc 4458  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 4729 ctggtctgcgttatacacaagctgttcgttccgtctgcacacgtgtgtatc 4788  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4459 ctggtctgcgttatacacaagctgttcgttccgtctgcacacgtgtgtatc 4518  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 4789 cacaagctctgtgtgaggaaagtgaggagacagagtgctcctgcggggagagctggcc 4848  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4519 cacaagctctgtgtgaggaaagtgaggagacagagtgctcctgcggggagagctggcc 4578  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 4849 tgcacatggccctgtgtctatcatagaggaggagcttaagaagaatctctctggaa 4908  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4579 tgcacatggccctgtgtctatcatagaggaggagcttaagaagaatctctctggaa 4638  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 4909 gctcatggcccatcatcatcttaagtattttaaattgattctgtctgcacagtgtc 4968  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4639 gctcatggcccatcatcatcttaagtattttaaattgattctgtctgcacagtgtc 4698  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 4969 aggaagggcacagaatgaggacagataaaccagatcatcttctgtctctctc -c 5027  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4639 aggaagggcacagaatgaggacagataaaccagatcatcttctgtctctctc -c 4758  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 5028 gtttttcaagtcagaggaaattgattgttctaaatgcatccacacgtcatgtctga 5087  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4759 gtttttcaagtcagaggaaattgattgttctaaatgcatccacacgtcatgtctga 4818  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 5088 ctgacgagtccacagctctcctaagttctgtcagatgaacataaagaagaaagaaac 5147  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4819 ctgacgagtccacagctctcctaagttctgtcagatgaacataaagaagaaagaaac 4878  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 5148 gttaacacatcacacagagaacatttcgggtccacagcgggtggtgaggaagctcact 5207  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4879 gttaacacatcacacagagaacatttcgggtccacagcgggtggtgaggaagctcact 4938  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 5208 ctgcggtcagatataagatgtgtgtgtggtgtcgggagatctcgtgtccatctcc 5267  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4939 ctgcggtcagatataagatgtgtgtgtgtgtgtcgggagatctcgtgtccatctcc 4998  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 5268 tctcatcttctgaacatcctgtatgttaaacacatggtcgggtgtctaagtgctgttaa 5327  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 4999 tctcatcttctgaacatcctgtgttaaacacatggtcgggtgtctaagtgctgttaa 5058  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 5328 tcccgatgtggaanaagctggaagtgaaagctcagacatcacatgattactttaaac 5387  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 5059 tcccgatgtggaanaagctggaagtgaaagctcagacatcacatgattactttaaac 5118  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 5388 agaaaaaagacatgtatgatagtctatttttttttatgtgacatttattttgt 5447  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 5119 agaaaaaagacatgtatgatagtctatttttttttatgtgacatttattttgt 5178  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 5448 gtgactgttttttaagaataatgtgttcacacacgtacccgtgtctcttcgtcattct 5507  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 5179 gtgactgttttttaagaataatgtgtgttcacacacgtacccgtgtctcttcgtcattct 5238  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 5508 gtgtcatgtgtctgttctcttaatacagtgctggcgggtgtcttaagtggtttacagtgtaac 5567  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 5239 gtgtcatgtgtctgttcttaatacagtcggcgggtgttaagtggtgttacagtgtaac 5298  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 5568 ggcacgtacaccttgatgacagtggtctcttctcaaacctccctggagctgtggaac 5627  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 5239 ggcacgtacaccttgatgacagtggtctcttctcaaacctccctggagctgtggaac 5358  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Oy 5628 agcttctctgtacataatgacactccttaataaagacatattctctctgt 5679  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 5359 agcttctctgtacataatgacactccttaataaagacatattctctctgt 5410  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 RESULT 2  
 AAH18587  
 ID AAH18587 standard; cDNA: 2718 BP.  
 XX  
 AC AAH18587;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:18774.  
 XX  
 KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 XX 28-JUL-2000; 2000EP-0116126.  
 PF  
 XX 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 PS  
 XX Claim 8; SEQ ID 18774; 2537bp + CD ROM; English.  
 PS  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination  
 CC of the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and  
 CC AAH13631 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95883 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification

	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883
--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

QY	1334	gcacgtgtacgacgtgtgtttaaacacacaaataatggcaagaagcttaatatccagctta	1399
Db	961	gcacccgttaagcagctgtgtttaaacacacaaataatggcaagaagcttaatatccagctta	1400
QY	1394	agaaagttacagaaggttttgggattccagctaccctcccaagatgttaacatagttgct	1453
Db	1021	agaaagttacagaaggttttgggattccagctaccctcccaagatgttaacatagttgct	1454
QY	1454	cagctccaacctrctgtgtgaaaaacattcccccggggggggccattcagatgtgcaaatgtgct	1513
Db	1081	cagctccaacctrctgtgtgaaaaacattcccccggggggggccattcagatgtgcaaatgtgct	1514
QY	1514	ttaagcagaagagacagactttaagagttaattagatagatttaagttggccaattcccaag	1573
Db	1141	ttaagcagaagagacagactttaagagttaattagatagatttaagttggccaattcccaag	1574
QY	1574	aggaagctgttcctgtctgttgagaagacccaagatggaaggaacgtgtgagctcttgct	1633
Db	1201	aggaagctgttcctgtctgttgagaagacccaagatggaaggaacgtgtgagctcttgct	1634
QY	1634	ttccgcagaagaagagccttccaccacaagggaacgtgaatgtgcagagccagccagatgcaga	1693
Db	1261	ttccgcagaagaagagccttccaccacaagggaacgtgaatgtgcagagccagccagatgcaga	1694
QY	1694	ttccaaaagaacagaaagcagaagaatgtggagattgttcttaacacctgattggcacaaggg	1753
Db	1321	ttccaaaagaacagaaagcagaagaatgtggagattgttcttaacacctgattggcacaaggg	1754
QY	1754	aattctgcacattgaatgaattccacatttaatgatccagatcctcagagcttggttgcaagt	1813
Db	1381	aattctgcacattgaatgaattccacatttaatgatccagatcctcagagcttggttgcaagt	1814
QY	1814	tccaaggttaaccgggtccaaaagagaacacacagagatttgggaattctgtcaagtlcattaa	1873
Db	1441	tccaaggttaaccgggtccaaaagagaacacacagagatttgggaattctgtcaagtlcattaa	1874
QY	1874	ttaatgtggagagcagcactctaaagatgggaagcttcgggtgaatgatccaactgtatgacg	1933
Db	1501	ttaatgtggagagcagcactctaaagatgggaagcttcgggtgaatgtatccaactgtatgacg	1934
QY	1934	taaatgtgaataaccctgttgggccaagaacaaacaaagatgcacatgggaacccctaaagaagt	1993
Db	1561	taaatgtgaataaccctgttgggccaagaacaaacaaagatgcacatgggaacccctaaagaagt	1994
QY	1994	ctatggtctacttgaagaagcctaataacgaggaatgatccagcttatcttttccaagggaataa	2053
Db	1621	ctatggtctacttgaagaagcctaataacgaggaatgatccagcttatcttttccaagggaataa	2054
QY	2054	gcaaggtgcataagctctgaagtcacacttggagcccccttgagacctgaagctggccattbaa	2113
Db	1681	gcaaggtgcataagctctgaagtcacacttggagcccccttgagacctgaagctggccattbaa	2114
QY	2114	cagcgtctgtgatgataaggaacgaagaatttcccatctccctatacagtgaggtatggagggc	2173
Db	1741	cagcgtctgtgatgataaggaacgaagaatttcccatctccctatacagtgaggtatggagggc	2174
QY	2174	ttgagtgatccccaagagaataatgtcctccctagatagataatgggt	2220
Db	1801	ttgagtgatccccaagagaataatgtcctccctagatagataatgggt	2221

RESULT 3  
 AAH78369  
 ID AAH78369 standard; DNA; 2718 BP.  
 XX  
 AC AAH78369;  
 XX  
 DT 26-NOV-2001 (first entry)  
 XX  
 DE Nucleotide sequence of a human protein kinase/protein phosphatase.  
 XX  
 KW Human; protein kinase; protein phosphatase; signal transduction; ss  
 XX

OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	CDS 33..2630
FT	/tag=a
FT	/product= "protein kinase/protein phosphatase"
XX	
PN	MO200109316-A1.
XX	
PD	08-FEB-2001.
XX	
XX	28-JUL-2000; 2000WO-JP05061.
PF	
PR	29-JUL-1999; 99JP-0248036.
PR	18-OCT-1999; 99US-0159590.
PR	11-JAN-2000; 2000JP-0118776.
PR	17-FEB-2000; 2000US-0183322.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;
PI	Senoo C, Nezu J;
DR	WPI: 2001-570286/64.
XX	
DR	P-PSDB; MAG67610.
XX	
PT	New genes encoding proteins with protein kinase/protein phosphatase
XX	activity, useful in the diagnosis and treatment of diseases -
PS	Claim 1; Page 67-76; 233pp; Japanese.
XX	
CC	The present sequence encodes a human protein kinase/protein phosphatase.
CC	It is expected that the protein kinase/protein phosphatase gene
CC	participates in signal transduction in cells. The protein kinase/protein
CC	phosphatase polypeptides and polynucleotides are useful for developing
CC	diagnostics and treatment agents for human and animal diseases. The
CC	protein kinase/protein phosphatase polypeptides are useful as target
CC	molecules in designing novel drugs. The protein kinase/protein
CC	phosphatase polynucleotides are useful as a source of probes and
CC	primers, which may be used to isolate homologous sequences.
SQ	Sequence 2718 BP; 840 A; 586 C; 688 G; 604 T; 0 other;
Query Match	29.8%; Score 1694; DB 22; Length 2718;
Best Local Similarity	99.8%; Pred. No. 0;
Matches 1844; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
OY	374 ttgaagttacacctcagtccttgcgagaataatccctcttaattgttcagcgagtgtg 433
Db	1 ttgaagttacacctcagtccttgcgagaataatccctcttaattgttcagcgagtgtg 60
OY	434 acccgctctaattggcctctccactctgtcagtgatagtaattttccttgaagac 493
Db	61 acccgctctaattggcctctccactctgtcagtgatagtaattttccttgaagac 120
OY	494 ctccaaggaaaaatcccacaacgctggtcaacaacagctggtctcctcaaagcaaacatg 553
Db	121 ctccaaggaaaaatcccacaacgctggtcaacaacagctggtctcctcaaagcaaacatg 180
OY	554 cttgggaagtcctaaaccctgcgcagcagaagaagaatgaanaactaacagaagccctccgcgg 613
Db	181 cttgggaagtcctaaaccctgcgcagcagaagaagaatgaanaactaacagaagccctccgcgg 240
OY	614 ataccagtaactggcgtctaaccaattccaagagacatatcgctcgtctctcgaagcca 673
Db	241 ataccagtaactggcgtctaaccaattccaagagacatatcgctcgtctctcgaagcca 300
OY	674 gtccacccaattggtggcgaagtgtgctggaagaacaagaatgaggaattggacagaag 733

[illegible]

OY 1814 tcaagtgtaacggtcacaagaagacacgcagagatttgggaattcttgaagtcacata 1873  
 Db 1441 tcaagtgtaacggtcacaagaagacacgcagagatttgggaattcttgaagtcacata 1873  
 OY 1874 ttaatggaagagacagatcctaagaatggaagcttgggtggaatgtaactgataagca 1500  
 Db 1501 ttaatggaagagacagatcctaagaatggaagcttgggtggaatgtaactgataagca 1500  
 OY 1934 taaatggagaatccctgttggcagaagacacacagatgcaatggaaacccataagaggt 1560  
 Db 1561 taaatggagaatccctgttggcagaagacacacagatgcaatggaaacccataagaggt 1560  
 OY 1994 ctatgtctactgaaagacataaaccggaagatgacagcttattgttgcagaagataa 2053  
 Db 1621 ctatgtctactgaaagacataaaccggaagatgacagcttattgttgcagaagataa 2053  
 OY 2054 gcaagtgcaatgagctggaagtcactggagccccctggagcctgagctgccccatgaaa 1680  
 Db 1681 gcaagtgcaatgagctggaagtcactggagccccctggagcctgagctgccccatgaaa 1680  
 OY 2114 cagcgttggatgataagagaacgaagaatttccattccctctacagtggtgagggagc 1740  
 Db 1741 cagcgttggatgataagagaacgaagaatttccattccctctacagtggtgagggagc 1740  
 OY 2174 ttgatgtaatcgccacgacgaagaatgctgcctcagtagaataatgggt 2220  
 Db 1801 ttgatgtaatcgccacgacgaagaatgctgcctcagtagaataatgggt 1847

RESULT 4  
 AAH72793  
 ID AAH72793 standard; cDNA; 1800 BP.  
 AC AAH72793;  
 DT 19-SEP-2001 (first entry)  
 DE Human cervical cancer marker nucleic acid 4067.  
 KM Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
 OS Homo sapiens.  
 PN WO200142467-A2.  
 PD 14-JUN-2001.  
 PE 08-DEC-2000; 2000MO-US33312.  
 PR 08-DEC-1999; 99US-0169681.  
 PR 21-DEC-1999; 99US-0171350.  
 PR 14-MAR-2000; 2000US-0189315.  
 PR 12-MAY-2000; 2000US-0203791.  
 PR 09-JUN-2000; 2000US-0210600.  
 PR 21-JUL-2000; 2000US-0220114.  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 PI Schlegel R, Deeds J, Berger A, Zhao X;  
 DR WPI; 2001-375006/39.  
 XX New isolated nucleic acid for diagnosing and treating cervical cancer  
 PT and for assessing and detecting compounds for treating the cancer -  
 PS Claim 1; Page 855-856; 1051IP; English.  
 CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
 CC polypeptides are useful to assess if a patient is afflicted with  
 CC cervical cancer or has a pre-malignant condition; to monitor the  
 CC progression of cervical cancer or a premalignant condition in a patient;

CC and to select and/or assess the efficacy of a compound or therapy for  
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
 CC useful for gene therapy.  
 XX  
 SQ Sequence 1800 BP; 578 A; 359 C; 466 G; 394 T; 3 other;

Query Match 17.6%; Score 1004; DB 22; Length 1800;  
 Best Local Similarity 99.4%; Pred. No. 0;  
 Matches 1604; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

OY 1462 atcatgtgaaataacattctcccccggggcgccctttagatgagtcgacctaagca 1521  
 Db 52 atcatgtgaaataacattctcccccggggcgccctttagatgagtcgacctaagca 1521  
 OY 1522 ggaagcagactatagagtaagtaagtagatttagtggcaaatcccaagagaagtc 1581  
 Db 112 ggaagcagactatagagtaagtaagtagatttagtggcaaatcccaagagaagtc 1581  
 OY 1582 gtctcgtgtgagaagcaccacagatggaaggaactgtgagccttcgtcttcgcag 1641  
 Db 172 gtctcgtgtgagaagcaccacagatggaaggaactgtgagccttcgtcttcgcag 1641  
 OY 1642 gaagacgcttccacccaagggaagtaatgcagaagcagaatgacagatcccaaa 1701  
 Db 232 gaagacgcttccacccaagggaagtaatgcagaagcagaatgacagatcccaaa 1701  
 OY 1702 gaagacgaagcagaatgagagatatttcttaaccctgagtgcacagaggaatttcg 1761  
 Db 292 gaagacgaagcagaatgagagatatttcttaaccctgagtgcacagaggaatttcg 1761  
 OY 1762 aatttgaagctccacataatgattcagatctgcagcgttcgtgtaagtccaagttcg 351  
 Db 352 aatttgaagctccacataatgattcagatctgcagcgttcgtgtaagtccaagttcg 351  
 OY 1822 aacggttaaaagaagacacgcagagatttgggaattcttgcagatccattatgaag 1881  
 Db 412 aacggttaaaagaagacacgcagagatttgggaattcttgcagatccattatgaag 1881  
 OY 1882 ggaagcagacttaagaatggaagcttgggtggaatgtaactgataagagaatgga 1941  
 Db 472 ggaagcagacttaagaatggaagcttgggtggaatgtaactgataagagaatgga 1941  
 OY 1942 gaatccctgttggcagaagacacacagatgcaatggaaacccataagaggtctatg 531  
 Db 532 gaatccctgttggcagaagacacacagatgcaatggaaacccataagaggtctatg 531  
 OY 2002 actgaagcacaataaagcagaatgatacagctattgttgcagaagataagcaagtc 2061  
 Db 592 actgaagcacaataaagcagaatgatacagctattgttgcagaagataagcaagtc 2061  
 OY 2062 aatgagctgaagtcacttggagcccccttggacactgagctgccccatgaaacgcgttg 2121  
 Db 652 aatgagctgaagtcacttggagcccccttggacactgagctgccccatgaaacgcgttg 2121  
 OY 2122 gatgatagaagaagaatttccattccctctacagttggtgagtgcaagataacagagtc 711  
 Db 712 gatgatagaagaagaatttccattccctctacagttggtgagtgcaagataacagagtc 711  
 OY 2182 tcggcagaagaatgctgcctcagtagatattgggtgagtgcaagataacagagtc 771  
 Db 772 tcggcagaagaatgctgcctcagtagatattgggtgagtgcaagataacagagtc 771  
 OY 2242 tccctacagtgaaatgcccccaagatgacatgtaattatagaagaagagtcgca 822  
 Db 823 tccctacagtgaaatgcccccaagatgacatgtaattatagaagaagagtcgca 822  
 OY 2302 gtgctctccacatctcttgcacagctctcttcacagctcccatgagatgagtgaggttc 882  
 Db 883 gtgctctccacatctcttgcacagctctcttcacagctcccatgagatgagtgaggttc 882  
 OY 2362 gtgagcagaatgctgtacttggcagaagctgcaatcagtgattcagcagagtcct 2421  
 Db 2421 gtgagcagaatgctgtacttggcagaagctgcaatcagtgattcagcagagtcct 2421









Db 545 tcagaatgtattagtagcagctgtaagcagtggtatatacaacacaaatagcgcaag 604  
QY 1377 gcttaataatcagcagcttaagaaggtacagaaggtttgggaattcagcatcactcaagaga 1436  
Db 605 gcttaataatcagcagcttaagaaggtacagaaggtttgggaattcagcatcactcaagaga 1436  
QY 1437 tgaacaataagtgctgaagctcaatctatgtgaaacattcccccggggggcgcc 1496  
Db 665 tgaacaataagtgctgaagctcaatctatgtgaaacattcccccggggggcgcc 1496  
QY 1497 catcaagatgagcagcagcttaagcagagacagacttatagagtaaatgagtag 1552  
Db 725 catcaagatgagcagcagcttaagcagagacagacttatagagtaaatgagtag 1552  
RESULT 7  
ID AAH33594 standard; cDNA; 823 BP.  
AC AAH33594;  
XX  
XX 03-SEP-2001 (first entry)  
DE Human colon cancer antigen encoding cDNA SEQ ID NO:650.  
XX  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX colorectal carcinoma; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200122920-A2.  
XX  
XX 05-APR-2001.  
XX  
XX 28-SEP-2000; 2000WO-US26524.  
XX  
XX 29-SEP-1999; 99US-0157137.  
XX 03-NOV-1999; 99US-0163280.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX WPI; 2001-235357/24.  
XX P-PSDB; AAG74163.  
XX  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers -  
XX Claim 1; Page 2675; 9803pp; English.  
XX  
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
XX cancer-associated nucleic acid molecules (N) and proteins (P), where  
XX the proteins are collectively known as colon cancer antigens. The colon  
XX cancer antigens have cytostatic activity and can be used in gene  
XX diagnosis and vaccine production. N and P may be used in the prevention,  
XX expression, for example, N and P may be used to treat disorders  
XX associated with decreased expression by rectifying mutations or deletions  
XX in a patient's genome that affect the activity of P by expressing  
XX inactive proteins or to supplement the activity of P by expressing  
XX by inserting the nucleic acids into a host cell and culturing the cell,  
XX and treatment of the proteins. N and P can be used in the prevention, diagnosis  
XX and AAB77789 represent sequences used in the exemplification of the  
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
XX missing at time of publication, meaning no sequences are present for  
XX SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
XX Sequence 823 BP; 249 A; 183 C; 201 G; 188 T; 2 other;

Query Match 12.7%; Score 725; DB 22; Length 823;  
Best Local Similarity 99.98; Pred. No. 0;  
Matches 775; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 777 tttagagcatataccgaactttctctgataatgataagctctagaagctccca 836  
Db 5 tttagagcatataccgaactttctctgataatgataagctctagaagctccca 836  
QY 837 cgaatgagggcctctgggaatccatgtatgctcttcaagtcgcagcgcgcaaacct 896  
Db 65 cgaatgagggcctctgggaatccatgtatgctcttcaagtcgcagcgcgcaaacct 896  
QY 897 ggggtatagtaagaagcattgggaagaggtgtgtaagcgtgaacatgaatcttttcg 124  
Db 125 ggggtatagtaagaagcattgggaagaggtgtgtaagcgtgaacatgaatcttttcg 956  
QY 957 tgaatgattgcatgtgcaggaatgaatgtgctgcagccttcgaatagaatttgaca 184  
Db 185 tgaatgattgcatgtgcaggaatgaatgtgctgcagccttcgaatagaatttgaca 184  
QY 1017 agcaacaatagtgttcgcaagcagtcgtacaaccatcattgtgtccatgtgtcc 1076  
Db 245 agcaacaatagtgttcgcaagcagtcgtacaaccatcattgtgtccatgtgtcc 1076  
QY 1077 tgcagcaataaagagcagatgacaacactaccgaagtgagaagaataactatc 304  
Db 305 tgcagcaataaagagcagatgacaacactaccgaagtgagaagaataactatc 304  
QY 1137 aagcgttttaagccttgacagcagcagatattgacaacagagtgtaacagtcagagct 1196  
Db 365 aagcgttttaagccttgacagcagcagatattgacaacagagtgtaacagtcagagct 1196  
QY 1197 tcaacagtgagagagagaccccgactgaaacaccccgctgagcagatagctcactc 424  
Db 425 tcaacagtgagagagagaccccgactgaaacaccccgctgagcagatagctcactc 424  
QY 1257 aagactactatagtcgacacacccctcggaacacacacacacacacacacacac 484  
Db 485 aagactactatagtcgacacacccctcggaacacacacacacacacacacacacacac 484  
QY 1317 tcaaatgtatttagtagcagctgtaagcagtggtataacacacacacacacacacac 544  
Db 545 tcaaatgtatttagtagcagctgtaagcagtggtataacacacacacacacacacacac 544  
QY 1377 gcttaataatcagcagcttaagaaggtacagaaggtttgggaattcagcatcactcaagaga 604  
Db 605 gcttaataatcagcagcttaagaaggtacagaaggtttgggaattcagcatcactcaagaga 604  
QY 1437 tgaacaataagtgctgaagctcaatctatgtgaaacattcccccggggggcgcc 664  
Db 665 tgaacaataagtgctgaagctcaatctatgtgaaacattcccccggggggcgcc 664  
QY 1497 catcaagatgagcagcagcttaagcagagacagacttatagagtaaatgagtag 1552  
Db 725 catcaagatgagcagcagcttaagcagagacagacttatagagtaaatgagtag 1552  
RESULT 8  
ID AAH07109 standard; cDNA; 556 BP.  
AC AAH07109;  
XX  
XX 26-JUN-2001 (first entry)  
DE Human cDNA clone (5'-primer) SEQ ID NO:3944.  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
XX Homo sapiens.  
XX  
XX EP1074617-A2.

```

XX 07-FEB-2001.
PD
XX 28-JUL-2000; 2000EP-0116126.
PF
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 1; SEQ ID 3944; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 556 BP; 154 A; 129 C; 144 G; 126 T; 3 other;
SQ

```

```

Query Match          7.2%; Score 407; DB 22; Length 556;
Best Local Similarity 99.6%; Pred. No. 4,9e-185;
Matches 507; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

DB 241 atactagtaactggtctcaaccatttcagagagacaaatgctgctgctctgagtgcca 300
OY 674 gtacccaatggtggcgaagtgtgagagaaacagagatgagatggagcgaag 733
DB 301 gtacccaatggtggcgaagtgtgagagaaacagagatgagatggagcgaag 360
OY 734 aggatatacagtcgtgttgaaacctgtgtgacatgctgacacgggtttggagcatatccca 793
DB 361 aggataacagtcgtgttgaaacctgtgtgacatgctgacacgggtttggagcatatccca 420
OY 794 actttctctgtagatgatagtgataaagctgtagaagctccccaagatgagggcctctg 853
DB 421 actttctctgtagatgatagtgataaagctgtagaagctccccaagatgagggcctctg 480
OY 854 gaatccatgtagtgccttcacgtgcga 882
DB 481 gaatccatgtagtgccttcacgtgcga 509

RESULT 9
AAF67583
ID AAF67583 standard; cDNA; 425 BP.
XX
XX AAF67583;
AC
XX
XX 09-APR-2001 (first entry)
DT
XX
XX Novel human polynucleotide, SEQ ID NO: 3345.
DE
XX
XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200102568-A2.
PN
XX
XX 11-JAN-2001.
PD
XX
XX 30-JUN-2000; 2000MO-US18374.
PF
XX
XX 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
XX (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klingner J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Cirenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones LM, Strache-Crain B;
XX
XX WPI: 2001-091805/10.
DR
XX
XX Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
XX Claim 9; Page 1045; 1046pp; English.
XX
XX The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,

```

CC dysplasia and hyperplasia.  
XX  
SQ Sequence 425 BP; 134 A; 86 C; 110 G; 95 T; 0 other;

Query Match  
Best local similarity 100.0%; Score 395; DB 22; Length 425;  
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1754 aattctgacattgaagtcaccacttaattgataagatcagatcagagccttggtcagt 1813  
DB 31 aattctgacattgaagtcaccacttaattgataagatcagatcagagccttggtcagt 1813  
QY 1814 tcaaaagtaaccggtcacaagaagaaccacgagatttggaattctgtcaagtcatta 1873  
DB 91 tcaaaagtaaccggtcacaagaagaaccacgagatttggaattctgtcaagtcatta 1873  
QY 1874 ttaatgagagagacgacatctaaagatggaagcctcggtgaatgatacactgatagcag 1933  
DB 151 ttaatgagagagacgacatctaaagatggaagcctcggtgaatgatacactgatagcag 1933  
QY 1934 taaatggaatccctgtgtggcagaacacgaagatgcatggaacccctaaagaagt 1993  
DB 211 taaatggaatccctgtgtggcagaacacgaagatgcatggaacccctaaagaagt 1993  
QY 1994 ctatgtctactgaagagcacaataacgaggaatgatacagctatttgtgaagagaataa 2053  
DB 271 ctatgtctactgaagagcacaataacgaggaatgatacagctatttgtgaagagaataa 2053  
QY 2054 gcaagtgaatgagctgaagtcacccctggagagcccccctggagcctgagcttgagaa 2113  
DB 331 gcaagtgaatgagctgaagtcacccctggagagcccccctggagcctgagcttgagaa 2113  
QY 2114 cagcgttgatgataagagaacgaagaattcccat 2148  
DB 391 cagcgttgatgataagagaacgaagaattcccat 425

## RESULT 10

AAH71480  
ID AAH71480 standard; cDNA; 597 BP.

XX AAH71480;

DT 19-SEP-2001 (first entry)

DE Human cervical cancer marker nucleic acid 2754.

KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

PN WO200142467-A2.

PD 14-JUN-2001.

PF 08-DEC-2000; 2000MO-US33312.

PR 08-DEC-1999; 99US-0159681.

PR 21-DEC-1999; 99US-0171350.

PR 14-MAR-2000; 2000US-0189315.

PR 12-MAY-2000; 2000US-0203791.

PR 09-JUN-2000; 2000US-0210660.

PR 21-JUL-2000; 2000US-0220114.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Deeds J, Berger A, Zhao X;

DR WPI; 2001-375006/39.

PT New isolated nucleic acid for diagnosing and treating cervical cancer  
and for assessing and detecting compounds for treating the cancer -

XX  
PS Claim 1; Page 566; 1051pp; English.

CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC progression of cervical cancer or has a pre-malignant condition; to monitor the  
CC and to select and/or assess the efficacy of a premalignant condition in a patient;  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.

SQ Sequence 597 BP; 180 A; 123 C; 156 G; 138 T; 0 other;

Query Match  
Best local similarity 100.0%; Score 269; DB 22; Length 597;  
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1402 acagaaggttggtgattcagatcactccagagatgtaacatagtggtcagctcca 1461  
DB 1 acagaaggttggtgattcagatcactccagagatgtaacatagtggtcagctcca 1461  
QY 1462 atctatggaataacatctcccccggggggcgccatcagatggtcagctcca 1521  
DB 61 atctatggaataacatctcccccggggggcgccatcagatggtcagctcca 1521  
QY 1522 ggaagacgactatagagtaaatgagtagattagtggcaaatcccaagagaagt 1581  
DB 121 ggaagacgactatagagtaaatgagtagattagtggcaaatcccaagagaagt 1581  
QY 1582 gtctcgtcttgagaagacccaagaatggaagaactgtgagcttctgtccgcaag 1641  
DB 181 gtctcgtcttgagaagacccaagaatggaagaactgtgagcttctgtccgcaag 1641  
QY 1642 gaagagcccttcacccaaggaagtga 1670  
DB 241 gaagagcccttcacccaaggaagtga 269

## RESULT 11

AAF66057  
ID AAF66057 standard; cDNA; 400 BP.

XX AAF66057;

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 1813.

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;  
KW breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

PN WO200102568-A2.

PD 11-JAN-2001.

PF 30-JUN-2000; 2000MO-US18374.

PR 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

PA (CHIR) CHIRON CORP.

PI (HYSE-) HYSEQ INC.

PI Williams LR, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Dimanac R;

PI Crkenjakov R, Dimanac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LW, Strache-Crain B;

DR WPI; 2001-091805/10.



```
XX Human lung tumour-specific cDNA #247.
DE
XX Lung tumour protein; lung cancer; cytostatic; vaccine; ss.
KW
XX Homo sapiens.
OS
XX WO200060077-A2.
PN
XX 12-OCT-2000.
PD
XX
XX 30-MAR-2000; 2000WO-US08560.
PF
XX
XX 02-APR-1999; 99US-0285323.
PR
XX 09-AUG-1999; 99US-0370838.
PR
XX 30-DEC-1999; 99US-0476235.
PR
XX 03-MAR-2000; 2000US-0518809.
PA
XX (CORI-) CORIXA CORP.
PI
XX Reed SG, Lodes MJ, Mohamath R, Secríst H;
PI
XX WPI; 2000-638466/61.
DR
XX
XX Novel lung tumor polypeptides and polynucleotides, useful for
PT detecting, monitoring or treating cancer, especially lung cancer -
PS
XX
XX Claim 3; Page 226-227; 243pp; English.
CC
XX The present sequence is given in a specification relating to compounds
CC for therapy and diagnosis of lung cancer. Polypeptides comprising at
CC least an immunogenic part of a lung tumour protein are disclosed.
CC The polypeptides are useful for inhibiting the development of cancer,
CC especially lung cancer. Samples of T cells expressing the polypeptides
CC may be used to inhibit the development of cancer. The polypeptides are
CC also useful for detecting and monitoring the progression of cancer,
CC especially lung cancer.
CC
XX Sequence 540 BP; 147 A; 123 C; 133 G; 136 T; 1 other;
SQ
Query Match 4.1%; Score 234; DB 21; Length 540;
Best Local Similarity 100.0%; Pred. No. 7.6e-102;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2227 ggttaataccagctgctccctacagtgatatagtccccaagatgacactgcatataga 2286
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 229 ggttaataccagctgctccctacagtgatatagtccccaagatgacactgcatataga 288
QY 2287 gatgacaggtgtgcagtgctctctccacatctctctgacacagtcctctccagctccat 2346
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 289 gatgacaggtgtgcagtgctctctccacatctctctgacacagtcctctccagctccat 348
QY 2347 gatgacaggtgtgttgacagcgacagatgctgtaacttgggccaagctgcaatcagtgat 2406
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 349 gatgacaggtgtgttgacagcgacagatgctgtaacttgggccaagctgcaatcagtgat 408
QY 2407 tcagcgacactgctcttgagtcacagatgctgacacagttcttgcttttaacga 2460
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 409 tcagcgacactgctcttgagtcacagatgctgacacagttcttgcttttaacga 462
RESULT 14
AAS68902
ID AAS68902 standard; cDNA; 540 BP.
XX
AC AAS68902;
XX
XX 26-FEB-2002 (first entry)
DT
XX Human lung tumour-specific SCC1-568 cDNA.
DE
XX Human lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW
```

```
KW antisense-therapy; vaccine; immune response; lung cancer; SCC1-568; ss.
XX
XX Homo sapiens.
OS
XX WO200172295-A2.
PN
XX
XX 04-OCT-2001.
PD
XX
XX 28-MAR-2001; 2001WO-US09991.
PF
XX
XX 29-MAR-2000; 2000US-0538037.
PR
XX 05-JUN-2000; 2000US-0588937.
PR
XX 18-AUG-2000; 2000US-0640878.
PR
XX 22-SEP-2000; 2000US-234517P.
PR
XX 01-NOV-2000; 2000US-0704512.
PR
XX 14-DEC-2000; 2000US-0738973.
PA
XX (CORI-) CORIXA CORP.
PI
XX Reed SG, Lodes MJ, Mohamath R, Secríst H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
PI
XX WPI; 2001-639201/73.
DR
XX
XX New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
PS
XX
XX Example 2; Page 276; 378pp; English.
CC
XX The invention relates to isolated lung tumour-specific proteins and
CC their corresponding cDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is human lung tumour-specific cDNA.
CC
XX Sequence 540 BP; 147 A; 123 C; 133 G; 136 T; 1 other;
SQ
Query Match 4.1%; Score 234; DB 23; Length 540;
Best Local Similarity 100.0%; Pred. No. 7.6e-102;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2227 ggttaataccagctgctccctacagtgatatagtccccaagatgacactgcatataga 2286
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 229 ggttaataccagctgctccctacagtgatatagtccccaagatgacactgcatataga 288
QY 2287 gatgacaggtgtgcagtgctctctccacatctctctgacacagtcctctccagctccat 2346
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 289 gatgacaggtgtgcagtgctctctccacatctctctgacacagtcctctccagctccat 348
QY 2347 gatgacaggtgtgttgacagcgacagatgctgtaacttgggccaagctgcaatcagtgat 2406
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 349 gatgacaggtgtgttgacagcgacagatgctgtaacttgggccaagctgcaatcagtgat 408
QY 2407 tcagcgacactgctcttgagtcacagatgctgacacagttcttgcttttaacga 2460
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 409 tcagcgacactgctcttgagtcacagatgctgacacagttcttgcttttaacga 462
RESULT 15
AAS68902
ID AAS68902 standard; cDNA; 510 BP.
XX
AC AAS68902;
XX
XX 13-FEB-2002 (first entry)
DT
XX DNA encoding novel human diagnostic protein #4706.
DE
XX
```



KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 XX MO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HXSE-) HXSEQ INC.  
 PA Drmanac RF, Liu C, Tang YT;  
 XX WPI: 2001-639362/73.  
 DR P-PSDB; ABG04715.  
 XX  
 PS New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 1; SEQ ID No 4706; 103pp; English.  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 510 BP; 133 A; 133 C; 144 G; 100 T; 0 other;

Query Match 3.7%; Score 211; DB 23; Length 510;  
 Best Local Similarity 99.2%; Pred. No. 8.7e-91;  
 Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 3320 ccattgctcacaagcgtacacctcagagccacagaaagggcatatgatgatgttcttgc 3379  
 DB 145 ccattgctcacaagcgtacacctcagagccacagaaagggcatatgatgatgttcttgc 204  
 QY 3380 atgcccagaatcaagaagcgcgggaattccaacacctcaacctgtagacaagtaacagatcaa 3439  
 DB 205 atgcccagaatcaagaagcgcgggaattccaacacctcaacctgtagacaagtaacagatcaa 264  
 QY 3440 ctctctagcaatcatgattcgatatacagcgtcttgaggaagaattcaggaagcaagaagaag 3499  
 DB 265 ctctctagcaatcatgattcgatatacagcgtcttgaggaagaattcaggaagcaagaagaag 324  
 QY 3500 atgaagaatgaagaatcgtcgcgcgaacctatagtttgagcaaccctgcccgaacgcac 3559  
 DB 325 atgaagaatgaagaatcgtcgcgcgaacctatagtttgagcaaccctgcccgaacgcac 384

QY 3560 ggcgcgcagacgcagagacgcgcgcacactcgtgtctcgttgagagtgacagatgcagcgcagc 3619  
 DB 385 ggcgcgcagacgcagatagcgcgcacactcgtgtcgttgagagtgacagatgcagcgcagc 444  
 QY 3620 ggcagagagagcgcgcgcagcgtccacagcagccagccagctcctcgcgcgc 3679  
 DB 445 ggcagagagagcgcgcgcagcgtccacagcagccagccagctcctcgcgcgc 504  
 QY 3680 aaag 3683  
 DB 505 aaag 508

Search completed: July 24, 2002, 14:29:07  
 Job time: 11709 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 09:16:33 ; Search time 6663.36 Seconds  
(without alignments)  
17866.522 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689  
Sequence: 1 atgaagtgaccgtgtgctt.....tcttcctgttaaaaaaaaaa 5689

Scoring table:  
OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Genembl:\*  
1: gb\_da:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pl:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	4848	85.2	5958	9	AF196185	AF196185 Homo sapi
2	4020	70.7	4071	9	AF467002	AF467002 Homo sapi
3	3411	60.0	4062	9	AF467003	AF467003 Homo sapi
4	3024	53.2	3960	9	AF467004	AF467004 Homo sapi
5	2577	45.3	2696	9	AK000761	AK000761 Homo sapi
6	2414	42.4	3470	9	AF196186	AF196186 Homo sapi
7	2003	35.2	196374	2	AL591464	AL591464 Homo sapi
8	1952	34.3	106198	9	AL160409	AL160409 Human DNA
9	1952	34.3	138688	2	AC025823	AC025823 Homo sapi
10	1899	33.4	2116	9	AK024668	AK024668 Homo sapi
11	1762	31.0	22279	2	U82210	U82210 Homo sapien
12	1694	29.8	2718	6	BD004905	BD004905 Novel gen
13	1694	29.8	2718	9	AK027735	AK027735 Homo sapi
14	1670	29.4	2967	9	AF332592	AF332592 Homo sapi
15	1670	29.4	3822	9	AF332593	AF332593 Homo sapi
16	1670	29.4	3933	9	AF467006	AF467006 Homo sapi
17	1497	26.3	1592	9	BC011711	BC011711 Homo sapi
18	1494	26.3	2004	9	AF177228	AF177228 Homo sapi
19	1452	24.3	3801	9	AF252293	AF252293 Homo sapi
20	958	16.8	3735	9	AF467005	AF467005 Homo sapi
21	714	12.6	1772	9	AK025892	AK025892 Homo sapi
22	427	7.5	427	6	AX072873	AX072873 Sequence
23	395	6.9	425	6	AX332939	AX332939 Sequence
24	387	6.8	104357	9	AL392123	AL392123 Human DNA
25	269	4.7	597	6	AX187060	AX187060 Sequence
26	252	4.4	400	6	AX071341	AX071341 Sequence
27	245	4.3	165146	2	AC011644	AC011644 Homo sapi
28	245	4.3	167163	2	AC013705	AC013705 Homo sapi
29	245	4.3	169337	2	AL356462	AL356462 Homo sapi
30	245	4.3	193131	9	AL390766	AL390766 Human DNA
31	234	4.1	540	6	AX321805	AX321805 Sequence
32	181	3.2	109816	9	AL360233	AL360233 Human DNA
33	181	3.2	138322	2	AC012241	AC012241 Homo sapi
34	181	3.2	196259	2	AL392171	AL392171 Homo sapi
35	134	2.4	305	6	AX187694	AX187694 Sequence
36	134	2.4	309	6	AX186627	AX186627 Sequence
37	121	2.1	173987	9	AL450337	AL450337 Human DNA
38	120	2.1	504	6	AX185634	AX185634 Sequence
39	115	2.0	167163	2	AC013705	AC013705 Homo sapi
40	110	1.9	339	6	AX184822	AX184822 Sequence
41	104	1.8	176405	9	AL138768	AL138768 Human DNA
42	96	1.7	5500	10	AB005549	AB005549 Rattus no
43	91	1.6	4167	10	AY026057	AY026057 Mus muscu
44	37	0.7	85713	9	HSY237C10_3	Continuation (4 of
45	37	0.7	168071	2	AL391479	AL391479 Homo sapi

#### ALIGNMENTS

RESULT	1	5958 bp	mRNA	linear	PRI 16-NOV-2001
LOCUS	AF196185				
DEFINITION	Homo sapiens atypical PKC isotype-specific interacting protein long				
ACCESSION	AF196185				
VERSION	AF196185.1	GI:13491609			
KEYWORDS					
SOURCE					
ORGANISM	human.				
REFERENCE					
AUTHORS	Fang,C.M. and Xu,Y.H.				
TITLE	Down-regulated expression of atypical PKC-binding domain deleted				
JOURNAL	asip isoforms in human hepatocellular carcinomas				
MEDLINE	Cell Res. 11 (3), 223-229 (2001)				
PUBMED	21520080				
REFERENCE	2 (bases 1 to 5958)				
AUTHORS	Fang,C. and Xu,Y.				
TITLE	Exon/Intron Structure and Splicing Variants of a Novel Human				
	Polarity Gene, HasIP				

QY	1	atgaagaatgacccggttgcttgcgaaagacccgggtggtcgtgctgcgcgtgcggagcgcac	60
Db	288	ATGAAAGTACCGTGTCTTTCGAGAGGACCCGGGGTGGTGTGGCTGTCCGGGAGCGGCAC	34.78
QY	61	atgaagttttcagcctcattccagcaaggcggltgacccgctacgcggaaagccaatcccaag	120
Db	348	ATGAAAGTTTTCAGCGCTTATCCAGAGGGCGGTGACCCGCTACGGGAAGGCATATGCCAAG	407
QY	121	gatccaacactcagatcacaggtgcacgcgtctggaataatggatgtgaagaaatactagac	180
Db	408	GATCCAAACTCTGGATTCAGGTGCATTCGCTTGGAAATATGGAAATGGAATACATGAC	467
QY	181	cttgatgacattctcttlygatlgtagcagacataaagacacagactbgtgacagtgttat	240
Db	468	CTTGATGACATTTCTTTGTGATGTGCGACAGACGATAAAGACAGACACGTGATGAGTGTGAT	527
QY	241	gagcagagatccacatacagcagaggtgtagtgcagccaggtccagttccacaggtaccagaag	300
Db	528	GAGCAGAGATCCACATCAACGAGAGGTGATGGCCACAGTCCAGTTTCCAGGGGTATCCCGAGAC	587
QY	301	ccagagatatttgtagtgcagcttgccacacaacatgltctcagcctttacagccttaccac	360

Db	588	CCAGAGATATTGGTATGTAGCTTGGCCACCAACAATGTCACGCTTTCACGCTTTACCA	64 7
Qy	361	gcaacaagltgaatttgaagtcacaccttaagtccttcaggacaataatgaccttcattgt	420
Db	648	GCACACAGTGAATATGAGTGCACACCTTAACTGATCTTCACGCAAAATATGCTCTTCAAGTT	707
Qy	421	cgagcgatgtatgacccagcttaatttgctctccactctctgtcatgtatgtaatttc	480
Db	708	CGAGCGAGTATGATACCCAGCTCTAAATTGGCCCTCTCCACTTCTGTCAAGTAAATTTT	767
Qy	481	tcctctgaagagccttcaaggaaaatacccaacgctgtgtatacaaaagcttgcttc	540
Db	768	TCTCTGTAAGAGCTTTCAGGAAAAATCCACACGCTTGCTGTAACAAACGCTGGCTTC	827
Qy	541	aagcaagaacactgtctggagctcctaaacctgcagagaaagaatgaactacaga	600
Db	828	AAGAGAAACCTGCTGGAGTCCCTAAACCTGCCACAGAAAGAAATGTAACACTACAG	887
Qy	601	agctcccgagatatactgaactcagctcagctcctaacaattttagagagcaatgctcgtg	660
Db	888	AGCTCCCGGGATCTATGTAATGCTCTTAAACAATTTCAAGAGACAATGCTCCTCG	947
Qy	661	tctctgagtgcagatcccaatggcttgggcagtgctgcgagaaacaagaacagatag	720
Db	948	TCTCTAGTCCAGTACCCCAATGCTGGGCAAGTGGCTGGAAACAGAACAGATGAG	1007
Qy	721	gattggacaagaagatataaagctgtgttgaacctgttgacatgctgacacggattg	780
Db	1008	GATGGACACGAAGAAGATATACAGTCTGTTTAACCTTGTGAATCTGCACAGGCTTGG	1067
Qy	781	gagcatatacccaactttctctgtgatatagtgtaagctctagaagtcaccaacagt	840
Db	1068	GAGCATATATCCCAACTTTCTCTGGATGATATGTTAAAGTCTGTAAAGTCCCAACGAT	1127
Qy	841	gagggggcccttgysaatccatgtagtgcttcctcaagtgctcagggcgysaaacctggg	900
Db	1128	GGAGGGCCCTCTGGGATCCATGATGTGCTTCCTTCACTGCTCGAGCGGCAAAACCCGGGG	1187
Qy	901	ttaattagtaaaacgattgagaaaggtgctgaagctgaaacatgaataacttttcgtag	960
Db	1188	TTATTATGTAACAAACGATTGGAAAGGTGTAAAGCTGTACATGTAAATTTTGTGTAG	1247
Qy	961	aattgttgcattgtccagattaatgatgtgcgacctcogaatatagaagatttgaacaaga	1020
Db	1248	AATGATTCCTATTGTCAGGATTTAAATGATGCGGACCTTCGAATATAGAAGATTGAACACGA	1307
Qy	1021	caacatagcttcgcgaagccaatgctgaaacccaataatttggttcattggttctctga	1080
Db	1308	CAACATATGTTTGGCCAAACCCATGCGTACACCCATATTTGGTCCATGTGGTTCTCCGA	1367
Qy	1081	gcaaataaagacgattgacaacatactccaagtgtgaagaagaacaaatctactatcaagc	1140
Db	1368	GCAAAATAAAGACAGTATGGAACAACATATCCCAAAAGGAGAAACAAATTAATTAACAC	1427
Qy	1141	cgatttagccctgcagacgagatattgtgcaacaagagtgtgaacagtgcagggctcac	1200
Db	1428	CGTTTAGCCCTGACACCCAGTATTTGACATACAGGAGTGTATACAGTGCAGGGCTTCAC	1487
Qy	1201	acggtgcagagagaccccgcaactgaaacccacccgctgagacatatagatctccactcaaga	1260
Db	1488	ACGGTGCAGAGAGACCCCGCACTGAAACCAACCGCGTGAAGCATATGACTTCACTCAAGA	1547
Qy	1261	ctactatattggcgaacccctcgggaaaaacacatccgctcagcctcggacccctcag	1320
Db	1548	CTACCTATATAGCGACACCCCTCTGGGAAAAACCATCTCCGCTCCAGCTCGGCACTCAG	1607
Qy	1321	aattgattagtagcagctgaaagcagtggtttaacaccaaaaaaataagcaagagctt	1380
Db	1608	AATGATTATTTAGTACGACTGTAAAGCACTGGTTATACACCAAAAAAATATGCAAGAGCTT	1667
Qy	1381	aataatcagcttaagaaggtacagaaggtttggattcagcatcactctccagatgctta	1440

Db 1668 AATATCCAGCTTTAAGAAAGTACAGAGGTTTGGATTTCAGCATCTTCCAGAGATGTA 1727  
Qy 1441 acaataagtggtctcaagctccaalctatgtgaaaaacatctcccccggggcgccatt 1500  
Db 1728 ACAATAGTGGGTGAGTCCCATCTATGTGAAAAACATTCTCCCGGGGGGGGCCANT 1787  
Qy 1501 caggaatggcgaactaaagcagagagacactatagagtaaatgagtagaattagtg 1560  
Db 1788 CAGGATGGCCCACTTAAAGCAGAGACAGACTTATAGAGTAAATGAGATAGATTAGTG 1847  
Qy 1561 ggcgaatcccaagaggaagtgttccgtgttggagaagacccaagaatggagaactgtg 1620  
Db 1848 GGCAAATCCCAAGAGGAAGTTGTTTGGCTGTGAGAGACACCAAAATGCAAGCAACTGTG 1907  
Qy 1621 agccttcgtctcttcgcaagaaagacgctccacccaaggaactgaatgcagacca 1680  
Db 1908 AGCCTTCTGCTTCTTGGCCAGGAAACGCCCTTCCACCCAAAGGAATGAAATCAGAGCA 1967  
Qy 1681 agccaagatgcgaatcccaaaagaaacgaagagaatgaggaatatgttcttaacct 1740  
Db 1968 AGCCAGATGCAGATTCCAAAAGAAAGCAAGAGATGAGATATTGTCTTACACT 2027  
Qy 1741 gatgacacaaggaattctgacatttgaagtcccaacttaagtatcagatctcagagc 1800  
Db 2028 GATGCAACAGGGAAATTTCTGACATTTGAAGTCCCATTAATGATTCAGGATCTCAGGC 2087  
Qy 1801 ctgtgtctcaagtgtcaaaagtaaccggtccaaagaacacagcagatttggaaactct 1860  
Db 2088 CTTGGTGTCAAGTCAAAAGTACCGGTCCTAAAGAAACACGACGATTTGGGAATCTT 2147  
Qy 1861 gtcaagatcatatataatgagagagcagcatctaaagatggaagcttcggtgtgaatg 1920  
Db 2148 GTCAAGTCACTTATTAAGAGAGAGACACATCTAAAGATGAAAGCTTCGGGTGATGAT 2207  
Qy 1921 caactgataagcaataatgaggaatccctgttggcgaagacaacaagaatgccatgaa 1980  
Db 2208 CAACGTAGACAGTAAATGAGAAATCCCTGTTGGCAAGCAACCAAGATCCCATGGAA 2267  
Qy 1981 accctaagaaggtctatgtctacttgaagcaataaagcaggaatgatcgaacttattgt 2040  
Db 2268 ACCCTAAGAAAGTGTATGTCTACTAAGCAATTAACGAGGAATATCATCAGTTATTGT 2327  
Qy 2041 gcaagagaataaagcaagtgcgaatgagctgaatgcacttggagaccccttggacctgag 2100  
Db 2328 GCAAGGAATAATAGCAATGCAATAGCTGAAGTCACTGGAGGCCCTTGACCTGAG 2387  
Qy 2101 ctgcccatttgaacacaggttggatgatatagaagaagaagaatttccattcccttaag 2160  
Db 2388 CTGCCCATTTGAACACGCGTTGGATGATAGAGAACGAAAGAAATTTCCCATTTCCCTTACAGT 2447  
Qy 2161 ggaattgaggggcttgaataatcgccagcagaataatgtgcctcactagatagaaatggt 2220  
Db 2448 GGGATTTGAGGGGCTTGATGAATCGGCCAGCAGAATGCTGCCCTCAGTAGATTAATG--- 2504  
Qy 2221 gagtcaagtaataacacagctgtccctcaagtgaatatgtgcccagaatgacactgtcatt 2280  
Db 2505 -----GGTAAATACCACTGTCCCTACAGTAAATATGCCCAAGATGACACTGTCAAT 2558  
Qy 2281 atagaagaatgacaggttgcaggtgtctctccacatctcttggaccagatccctccagc 2340  
Db 2559 ATAGAAAGTACAGGTTTCCAGTGTCTTCTCCACATCTCTTGACACAGTCCCTTTCACGC 2618  
Qy 2341 tcccatgtatgttggggttctgtgacgagatgctgtglaacttggccaagctgcaatc 2400  
Db 2619 TCCCATGTATGATGTGGGTTTGTGACGGCAGATGCTGTACTGTGGCCAAAGCTCAGATC 2678  
Qy 2401 agtgaattcaagcagactgtctcttgaatccagatgttgaatccagttcttgcattcaaga 2460  
Db 2679 AGTGAATTAGCCGACTGCTCTTGTGATCCAGATGTTGATCCAGTTCTTCTTCAACGA 2738  
Qy 2461 gaagaatttgaagctcagaagatagtccagaaaaacgacaaaagcaatttccagatgcagt 2520  
Db 2739 GAAGATTGGACGTCAAGATATGTCAAGAAACGCAAAAGCAATTTTCAGATCCAGT 2798

Qy 2521 caattggatttctgttaaaaaacagaaatacaaaaagcatgtatttaagtatactgacag 2580  
Db 2799 CAATTGGATTGTGTTAAACACAGAAATCAAAAAGCATGATTAGTATAGCTGACGAG 2858  
Qy 2581 actaaactcaatacagttgatagaaccaagaagagttctcccaagagatgttggctct 2640  
Db 2859 ACTAAATCCAAATACAGTGTGATCACGAAAGCAGGTTCTCCACGACGAGATGTGGTCCCT 2918  
Qy 2641 tccctgggtcttaagaagaagctcgttgggaggtctgcacaacgcaatttccgaggtg 2700  
Db 2919 TCCCTGGGCTTAAGAAAGTCAAGCTTATTGAGAGTCTCCACACCGCATTTCCCGAGGTG 2978  
Qy 2701 acttgaatgggataatcttcttccatcgctccacggtcccgagataatcagagcaagga 2760  
Db 2979 ACTTTGAATGGGGATTTCTCTTCCATCGCTCCACGGCCCGGATATATCAGAGCAGGGGA 3038  
Qy 2761 tgcgaatgagactcgaagctgcacatcgacaatcttatagtataaacccggtatgat 2820  
Db 3039 TCCCAATGACAGCTTCAGACCTCCATCGACAAATCTTATGATTAACCCCGGTAGATGAT 3098  
Qy 2821 gatgatgaaggaatggaagccttggagaagaacagaaagaattcaagaatcagaagaga 2880  
Db 3099 GATGATGAAGGCAATGGAGACTTTGGAGAGACACAGAAAGATTCAAGATCAGGGAGA 3158  
Qy 2881 gagtctgataccaagcagatgatacagcttccactctctgtgagagacaaatgaaatgga 2940  
Db 3159 GAGTGTGTCCACAGCCAGTATCAGCTTCCCATCTCTGTGAGAGACAAATGATGGA 3218  
Qy 2941 aaccaagaagaagtgtataagactgataaaaaaagaataaaaactggaagaagaagaag 3000  
Db 3219 AACCAAGAAAGGTGATGAAGACTGTAGAAAAAAGATTAACCTGCAAAAGCAAAAGAG 3278  
Qy 3001 aagaatagaagataaagagaagataaaatgaaagccaagaagaagatgcttgaaggtctg 3060  
Db 3279 AAGATAGAGATAGAGAGAGATTAATGAAAGCCAAAGAGGAATGCTGAAGGGGCTTG 3338  
Qy 3061 ggaagacatgttcaaggttcttgcacaacatcgaaaaagatgaaagaatltagaanaacggttaa 3120  
Db 3339 GGAGACATGTTCAAGTGTGGCAAAACATCGAAAAAGATGACAAAGTTGAAAAACGGGTAAA 3398  
Qy 3121 ataaaaatacagaagatccttatacagaagaagagagaatacgaatgaagcagaagcag 3180  
Db 3399 ATTAATAATACAGGAATCCCTTACATCAGAAAGAGAGAGATACGAATGAACGAGCAG 3458  
Qy 3181 ggaagattccaagccaaactcgaagaatttaggaacggaagctgcgaagcgttgactat 3240  
Db 3459 GAGAGATTCAAGCCAAACCTCGAGAAATTTAGGAAACGACAAAGCTCGAGGCTGACTAT 3518  
Qy 3241 gctgaatccaagatttcaatcggaacatttggctgtgaatgagttaaattgataaggagga 3300  
Db 3519 GCTGAATAATCAAGATTTTCATCGAGACATTTGGCTGTGATGATGATTAATGATGGGGGA 3578  
Qy 3301 gttcttcttaagaaggttccatgtctccaagttagacactcagaagcccaagaagaag 3360  
Db 3579 GTTTCTCTTATGAAGGTTCCATGTGCTCAACGCTAGACCTCAGAGCCACAGAGAGGG 3638  
Qy 3361 catatgatgatgtcttgtatgcccgaagtcaagaagccgaggaatttccaaacctcaact 3420  
Db 3639 CATATGATGAGTGTCTTTGTATGCCCCAAGTCAAGAAACCCGGGAATTCAAACCTTACCT 3698  
Qy 3421 gtagacagtaacagatccaactcctagcaatctatgatacagatgcgttgaagcaaga 3480  
Db 3699 GTAGACAGTAACAGATCAACTCTCTAGCATCATGATCGGATACAGGCTGAGGCAAGAA 3758  
Qy 3481 ttccagcaagaagaagaatgaagaatgtagaagaatcgtccgaggaactatagtttgag 3540  
Db 3759 TTTTCAGCAAGCAAAACCAAGATGAAGATGTAGAAAGCTGTGGCGGACCTATGATTTGAG 3818  
Qy 3541 caaccctggcgaagcagagcgcgagcagagagagggcggaactcgggtgtccgtggag 3600  
Db 3819 CAACCTTGCCCAAGCAGCGCCGCCGACGACAGAGCGGGCGCACTTCGGTGTCCGTGGAG 3878

[illegible][illegible]

ACCESSION AF467002 GI:18568343  
VERSION AF467002.1  
KEYWORDS  
SOURCE  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 4071)  
AUTHORS Gao, L., Macara, I.G. and Jobery, G.  
TITLE Multiple splice variants of PAR3 and of a novel related gene, PAR3L, produce functionally different proteins  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4071)  
AUTHORS Gao, L., Macara, I.G. and Jobery, G.  
TITLE Direct Submission  
JOURNAL Submitted (10-JAN-2002) Center for Cell Signaling, University of Virginia, PO Box 800577, Charlottesville, VA 22908, USA  
FEATURES  
source location=Qualifiers  
1..4071 /organism="Homo sapiens"  
gene /db\_xref="taxon:9606"  
1..4071  
/gene="PAR3"  
/note="PAR3"  
1..4071  
/gene="PAR3"  
/note="PAR3: multi-PDZ domain containing protein; alternatively spliced"  
/codon\_start=1  
/product="partitioning-defective 3 protein splice variant a"  
protein\_id="AAL76042.1"  
/db\_xref="GI:18568344"  
translation="MKVTVCGRTRRVVPCGDGMKVFSLIQAVTRKAIADPNY  
WIOYHREHGDGILLDLIDLCYVADDDKRLVAAPDODPHHGGDPSASSTGOSPE  
IRGSELCTNNVSARPOYATSEIPEVPSVLARNPLRYRSDPALLIGLSISYSDSNF  
SSEPRKKNPTRMSTAGFLKONTAGSPKTDKDKDEYRSLPHDTSWMSQFORDNA  
RSSLSASHPHWKMLEKQEDDEDEEDNSVEVEGHADGLEHLPNSLDDMKLVE  
VPNDGSLPIHVVESFARGRTGLVLRLEKGAHEHNFRENDCLVRINDDLRN  
RREOAOHMFORARATPPIIMFHYVPAANKEDQEQLOSSEKNYSSRSPSOYIDNR  
SVNSAGLHTVORAPRLNHPEDQIDSHSLRPSHAPSGKPPAPASAPONVSTVSSG  
YNTKKIGKRLNIOCLKGTEGIGFSTSRDVTIGSAPVYVNTLPRGAIDODGLKAC  
DLRLVNGVYLCKSQEVEVSLRSTMGVSLVLRQEDAFPRELNAEPQOMQIP  
KETAEDEDIVLPDGTREFLPTEFEVPLNDSGSLGVSVKNRSEKEMHADGIVKSI  
INGGAASKDRRLVNDQLIAVNGESLGTQMDMELLRSMSTEGKGMIOUIVAR  
RISKCNELKSPSPGPELPIETALDRERISHSIXEIGLSDSPSRNALSRIMG  
ESGKYLOSPVYNNPDDVTIIEDRLPVLPHLSIDSSSSHDVGFPTADAGWAKA  
AISDASCSLSPVDVPLAFORRGROSGSEKRTKPSDASQLDPYTKRSKMDLG  
IADERTLNTYDDOKAGSPRDVPSLGLKSSLSLSIQTAAYETLNDIPEHPRK  
IRRGCNESFRAIDKSYDKPAVDDEDEGETLEETESSRGRRESVSTASDQPSH  
SLEROMNGNEGKTDKDKDKTEKREKDKDEKDKAKKMLKGLGDFRGRKHK  
KDKLIEKTKIKIIESFTSEERIIMKOEORIOAKREFREPAREDEYIEDFHR  
TFGDDDELMTGGVSYEGSMALNARPOSREGHMDLVAVKRPNRKSPVDSNRS  
TPSNHDIORLROFOAKODEVEDERARYSEOPBNAPAPQSGHSHSVYQMO  
RORERESSQOAKROTSLPROSRRKNAASVSQSDMONTSPGSGFQSAKKNPYSSY  
QGSNGTGLGGHGFARVMLETQELRLQROERKEEQOMKQPPSEGPSNDSYKQVDP  
YAPKGPFRDVPSPSOVARLNLQTPEKGRPYs"  
814..1074  
/gene="PAR3"  
/note="Region: PDZ domain"  
1378..1638  
/gene="PAR3"  
/note="Region: PDZ domain"  
1765..2028  
/gene="PAR3"  
/note="Region: PDZ domain"  
BASE COUNT 1220 a 923 c 1079 g 849 t  
ORIGIN  
Query Match 70.7%; Score 4020; DB 9; Length 4071;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgaaagtacccgtgtcttcggaacggaacccgggtgtgtcgtgcggtgagcgccac 60  
DB 1 ATGAAAGTACCCGTGTGCTTCGAGAGGACCCGGGTGTGTGCGCGGAGCGGCAC 60  
QY 61 atgaaagtcttaagcctcatccagcagcggtgtgacccgttacggaaggccatgcgaag 120  
DB 61 ATGAAAGTCTTAAGCCTCATCCAGCAGCGGGTGTGACCCGCTACCGGACCATCGCAAG 120  
QY 121 gatccaactactgtgatacaaggtgtcgccttggaaacagagagatgagaaatactaac 180  
DB 121 GATCCAACACTACTGTGATACAGGTGCATCCCTTGGAAACATGAGATGAGCAATACTAGAC 180  
QY 181 ctgtatgacattcttctgtatgtatgacagacgaataaagacagactggttagcaagtgttat 240  
DB 181 CTGTATGACATTCTTCTGTATGTATGACAGACGATAAAGACAGACTGTTAGCAGTTGAT 240  
QY 241 ggcgcagatgccacatcaacgagatgtatgtgcacacagtgtccagtccaggtaccagagc 300  
DB 241 GGCAGGATGCCACATCACAGAGGTGATGTGCACACAGTGCACAGTCCACGGGTACCCAGAC 300  
QY 301 ccagagatatctgtgtatgtagcttggcaccacaaatgctcagccttacgacctaacaa 360  
DB 301 CCAGAGATATTTGTGTATGTAGCTTGGCACCAACAATGTCTCAGCCTTACGCTTACCAA 360  
QY 361 gcaacaagtgaatitgaaggttcaacacctcagtccttcgacaataatgctcttcagt 420  
DB 361 GCAACAAGTGAATITGAGGTTCACACCTTCAGTCTTCGACAAATATGCTCTTATGTT 420  
QY 421 cgaagcagtagtgaacccagcttaattgtgctctccaattctgtcagtgatagtaattt 480  
DB 421 CGAAGCAGTAGTGAACCCAGCTTAATTGTGCTCTCCAACTTGTCAAGTATGTAATTTT 480  
QY 481 tctcttgaagagccttcaagagaaataatccacacagctgtgtcaacacagctgctcc 540  
DB 481 TCTCTTGAAGAGCCTTCAAGAAATAATCCACACGCTGTGTCAACACACTGCTTCTTC 540  
QY 541 aagcgaacactgtctggaggtccttaaaacctcgaaaggaagaagaatgaaactacaga 600  
DB 541 AAGCGAACACTGTCTGGAGGTCTCTAAACCTCGAAGGAAAGATGAATAACTACAGA 600  
QY 541 aagcgaacactgtctggaggtccttaaaacctcgaaaggaagaagaatgaaactacaga 600  
DB 541 AAGCGAACACTGTCTGGAGGTCTCTAAACCTCGAAGGAAAGATGAATAACTACAGA 600  
QY 601 agccctccgaggtactactgtactgtgttaacaaattccagagagaaatgtcgtctg 660  
DB 601 AGCCCTCCGAGGTACTACTGTACTGTGTTAACAAATTCAGAGAACATGCTCTCCTCG 660  
QY 601 accctcccgaggtactactgtactgtgttaacaaattccagagagaaatgtcgtctg 660  
DB 601 ACCCTCCGAGGTACTACTGTACTGTGTTAACAAATTCAGAGAACATGCTCTCCTCG 660  
QY 661 tctctgaagtgccagtcacccaatgtgtggaagtggtcggagaaacagagatgag 720  
DB 661 TCTCTGAAGTGCCAGTCAACCAATGTGTGGAAGTGTGCGGCAAGTGGCTGAGAACACAGATGAG 720  
QY 721 gatggagacgaagagatatacagtcgttgaacctgttgcacatgtctgacacgggtttg 780  
DB 721 GATGGAGACGAAGAGATATAACAGTGTGTTGAACCTGTTGACATGCTCAGACGGGTTTG 780  
QY 781 gagcatatacccaacttctctgtatgtatgttaagtcgttagaagtccecaagat 840  
DB 781 GAGCATATACCCAACTTCTCTGTATGTATGTAAACCTGTATGAAGTCCCAACGAT 840  
QY 841 ggaagagcctctggaaatccatgtagtgccttcagtgctcgaagcgagaaacccctggag 900  
DB 841 GGAAGAGCCTCTGGAAATCCATGTAGTGCCTTTCAGTGTGAGGCGGCAACCCCTGGAG 900  
QY 901 ttattagtaaaacgaattggagaaaggtgtgaaagcttgaacatgaaatcttttcgtgag 960  
DB 901 TTATTAGTAAACGATTTGAGAAAGGTGTAAAGCTGAACATGAAATCTTTTTCGTGAG 960  
QY 961 aatgattgcatgtcgaagataatgattggtgcagacctcgaataagaagaatttaacaagca 1020  
DB 961 AATGATTGCAATGTGAGGATTAATGATGTGCGACCTTCGAATAAGAAATTAAACAAAGCA 1020  
QY 1021 caacatatttgcgcaaacgcatgtacacccatcatattggttccatgtgttctctga 1080  
DB 1021 CAACATATTGTCGCAACGCAATGCTTACACCCATCATATTGTTCCATGTGTTCTCTGCA 1080

```
OY 1081 gcaataaagagcagatgtaacaaactaaccaaagtgaagaagacattactattcaagc 1140
Db 1081 GCAATAAAGAGCAGATGTGAACAACATATCCAAAGTAGAAGAAACATTAATATTCACAC 1140
OY 1141 cgttttagccctgacagccagctatattgcaaaagaagtgtgaacagtgcaaggtctaac 1200
Db 1141 CGTTTtagccctgacagccagctatattgcaaaagaagtgtgaacagtgcaaggtctaac 1200
OY 1201 accgttgagaagacaccccgactgaacacaccccgctgaagcagatagactctcactcaaga 1260
Db 1201 ACCGTTGAGAAGACACCCCGACTGAACACACCCGCTGAAGCAGATAGACTCTCCTCAAGA 1260
OY 1261 ctacctatagcgacacacccctcggaataaacacacatccgctccagcctcgcaacctcag 1320
Db 1261 CTACCTATAGCGACACACCCCTCGGAATAAACACATCCGCTCCAGCTCCGCAACCTCAG 1320
OY 1321 aatgtaattagtaagcagtgtaagcagtggtatatacaccacaaataaagaagaagcct 1380
Db 1321 AATGTAATTAGTACAGCTGTATAGCAGTGTATACACCAAAAATAAGCAGAGAGCTT 1380
OY 1381 aatatacagcttaagaagaagtaacagaagtttggaattcagcattccactctcagagatgta 1440
Db 1381 AATATCCAGCTTAAGAAGAAGTACAGAAAGTTGGGATTGAGCATCACTTCAGAGATGTA 1440
OY 1441 acaataagtgagctcagctccaatctatgtgaataaacaattcccccgaggagcgccact 1500
Db 1441 ACAATAAGTGAGCTCAGCTCCAATCTATGTGTAAGAAACATTTCTCCCGGGGGGGCGCAAT 1500
OY 1501 caggatggcagacttaagaagcagagagacagacttataaggtlaaatggaattagttagtg 1560
Db 1501 CAGGATGGCCAGCTTAAGAAGCAGAGACAGACTTATAGAGTAATGAGATGAGATTAGTG 1560
OY 1561 ggcacaatcccaagaagaagtgttctgcgtgtgaagaagacaaagtgtgaagaagactgtg 1620
Db 1561 GGCACAATCCCAAGAGAAAGTTGTTGCGCTGTGAGAAAGCAAGAGTGAAGAAAGAACTGTG 1620
OY 1621 agccttcgtctcttcgcgcgcgcgcgcgcgcctccaccgaaggaaactgaatcagagcca 1680
Db 1621 AGCCTTCTGCTCTTTCGCCGAGAGAGAGCTTCCACCAGGAAGAACTGAATGACAGAGCCA 1680
OY 1681 agccagatgcagattcccaaaagaagaagaagcagaagatgaagatattgtcttcaacct 1740
Db 1681 AGCCAGATGCAGATTCCAAAGAAAGAAAGCAAGAGATGAGATATTGTTTACACT 1740
OY 1741 gatgagcagcaggaattcttgaacatttgaagtcccaattatgaattcagagtcgcagcgc 1800
Db 1741 GATGAGCACAGGGAATTTCTTGACATTGAAGTCCCACTTAATGATTCAGATTCGACAGC 1800
OY 1801 cttgtgtcagtgtcaaaagttaaccggtcaaaagaagaacccgcagatttgggaactctt 1860
Db 1801 CTTGTGTCAAGTTCMAAAGTAAACGGGTCAAAAGAAACCAAGCAAGATTTGGGAATCTTT 1860
OY 1861 gtcaagtcattatgaatgagagcagcagcatctaagaatgaaggtctcggtgtgaatgat 1920
Db 1861 GTCAAGTCAATTAATGAGAGAGCAGCATTAACATGGAAGCGCTTGGCGGTGAATGAT 1920
OY 1921 caactgataagcagtaaatgagaaatccctgtgtgcaagaacaaacaaatgcatgaa 1980
Db 1921 CAACGTAGAGCAGTAATGAGAAATCCCTGTGGCAAGCAAAACCAAGATGCCATGGA 1980
OY 1981 accctaagaaggtctactgtactatgaagcaataaagcagaatgaatccagcttaattgt 2040
Db 1981 ACCCTAAGAAGGTCTATGTCTACTGAAGCAATAAAGAGAAATGATTCAGCTTATTTGT 2040
OY 2041 gcaaggaagaataagcaagtgcataagctgaagtcacccctggagagcccccctggacctag 2100
Db 2041 GCAAGGAAATAAGCAAGTGCATGAGCTGAGTCACTCGGAGAGCCCCCTGGACCTCAG 2100
OY 2101 ctgcccactgaaacagcagcttgatagatagaagaagaagaatttccacttccctctacagt 2160
Db 2101 CTGCCCACTGAACAGCAGCTTGATGATAGAGAAAGAAATTTCCCATCTCTACAGT 2160
OY 2161 ggagattgaggagcttgatgaatcgcacagaaatgtgcccctcagtaggataatgagt 2220
Db 2161 GGAGATTGAGGAGCTTGATGAATCGCACGAATGTGCCCCCTCAGTAGGATAATGAGT 2220
OY 2221 gaatcagtaataacacagctgtccctacagatgaatagccccagaatgacatgtaalt 2280
Db 2221 GAATCAGTAATAACACAGCTGTCCCTACAGATGAATAGCCCCAAGATGACACTGTCAAT 2280
OY 2281 atagaagatgacaaggttgcagctgcttcccaatcctctctgaacagcttccctccagc 2340
Db 2281 ATAGAAGATGACAAGGTGCGAGTGTCTCTCCACATCTCTGACAGCTCTCTCCAGC 2340
OY 2341 tcccatgatagtgtggttgtgtacggcagatgtcgtactgtggcgaagctgtaac 2400
Db 2341 TCCCATGATGATGTGGGTTGTGTACGGCAGATGTCTGTACTGTGGGCAAGCTCGCAATC 2400
OY 2401 agtgattacgcagactgctcttgatccagatgtatagctcagttcttcttcaaga 2460
Db 2401 AGTGATTACGCAGACTGCTCTTGATCCAGATGTGATTCAGTTCTTCTGTTTCAAGCA 2460
OY 2461 gaagatttgaagctgaagatgatagtcaagaataacgcacaagaatcttcagatgcagt 2520
Db 2461 GAAGATTTGAAGCTGAGATGATGTCAAGAAACGCACAAGCAATTTCAATGCGCAAT 2520
OY 2521 caattggaattcgttaaaacagcaaaatcaaaagcatgattgtatagctgaagc 2580
Db 2521 CAATTGGATTTCGTTAAACACGAAATCAAAAAGCATGATTTAGGTATACCTGACAG 2580
OY 2581 actaaatcaatacagtgatgatagcagaaagaagttctccagcagagatgtggtcct 2640
Db 2581 ACTAAATCAATACATGATGATGATGACAGAAAGCAGAGTTCTCCAGCAGAGATGTGGTCT 2640
OY 2641 tccctggtctgaagaagtcagctcgtgtgagagtcgcgcagacgcgcgaggtgcccagtg 2700
Db 2641 TCCCTGGTCTGAAGAAGTCAGCTCGTGGAGATTCGACACGCGAGTTGCGAGGTG 2700
OY 2701 acttgaatgggatacttcttcacatgctcagccgcgcgcgcgcgcgcgcgcgcgcgcgc 2760
Db 2701 ACTTGAATGGGATATTCCTTCTTCATGCTCCAGCGCGCGATTAATCAAGCAGGGA 2760
OY 2761 tgcataagaaagcttcaagctgacatgcacaaactctatgataaaccgcgcgcgcgcgcgc 2820
Db 2761 TGCATAGAAAGCTTTCAGAGCTGCCATCGACCAATCTTATGATTAACCCGCGGTAGATGAT 2820
OY 2821 gatgaatgaagcagatgaagccttgaagaagacacagaagaagttccaagatcagagaga 2880
Db 2821 GATGATGAAGCAGATGAGACCTTGGAAGAGACACAGAAAGATTCAAGATCAGGAGA 2880
OY 2881 gagtgtatccacagccagtgatcagccttcccaactctctgagagaagaatgatatga 2940
Db 2881 GAGTGTATCCACAGCCAGTGAATCACTTCCACCTCTGAGAGACCAATGATGGA 2940
OY 2941 aaccaagaagaagtgataagcttgataagaaagaataaactggaataaagaagaagaag 3000
Db 2941 AACCAAGAAAGTGATGATGACTGTGATGAGAAAGAAAGATTAACGGAAGAAAGAAAG 3000
OY 3001 aaagatgaataaagaagaagaataaatagaagaagcagaagaaggaatctgaagagctg 3060
Db 3001 AAAGATGAATAAAGAGAGATTAATAATGAAGCAAGAAAGGAATCTGAAAGGCTTG 3060
OY 3061 ggaagacatgtcaggttttgcaaacactgaaagaatgaacagatgagaagaacggttaaa 3120
Db 3061 GGAGACATGTTCAAGTTTGGCAAACTCGAAAGATGCAAGATTTGAGAAACCGGTAAA 3120
OY 3121 ataaataacagaagaatccttatacagaagaagaagaatgaagaatgagaagaacggttaa 3180
Db 3121 ATAAATAACAGAAATCTTTTACATCAGAAAGAGAGAGATACCAATGAAGCCAGGACAG 3180
OY 3181 gagagatccaagccaacactgagaatattgagaacgaagctcgaagcgttgactat 3240
Db 3181 GAGAGATTCAAAGCCAAACTCGAATATTAGGGAACGAACGTCGAGAGCTGACTAT 3240
OY 3241 gctgaatccaagatttcacatgcacatttgctgtatgataatgatatggtgga 3300
Db 3241 GCTGAATCCAAGATTTCACTGCACATTGCTGTATGATAATGATATGTTGGTGA 3300
```





OY 181 ctgtatgacattcttctgtatgtacagacataaagacagactgttagcagtttcat 240  
 Db 181 ctttgatgacattcttctgtatgtacagacataaagacagactgttagcagtttcat 240  
 OY 241 gagcagagatccacacacacacagaggttgatgacacagttccacaggtttagc 240  
 Db 241 gagcagagatccacacacacacagaggttgatgacacagttccacaggtttagc 240  
 OY 301 ccagagatattctgtatgtatgactgtgacacacacacacacacacacacacacac 300  
 Db 301 ccagagatattctgtatgtatgactgtgacacacacacacacacacacacacacac 300  
 OY 361 gcaacaagtgaaattgaggttgacacacacacacacacacacacacacacacacac 360  
 Db 361 gcaacaagtgaaattgaggttgacacacacacacacacacacacacacacacacac 360  
 OY 421 cgaacagtagtgac 420  
 Db 421 cgaacagtagtgac 420  
 OY 481 tcccttgaagagccttcaagaaataccacacacacacacacacacacacacacacac 480  
 Db 481 tcccttgaagagccttcaagaaataccacacacacacacacacacacacacacacac 480  
 OY 541 aagaagaacactgtctgaggttccataaacctgcagacagagaagaataaactaacaga 540  
 Db 541 aagaagaacactgtctgaggttccataaacctgcagacagagaagaataaactaacaga 540  
 OY 601 agcctccgcgggatactagtaactgttcaacacacacacacacacacacacacacac 600  
 Db 601 agcctccgcgggatactagtaactgttcaacacacacacacacacacacacacacac 600  
 OY 661 tctctgaatgcagtcac 660  
 Db 661 tctctgaatgcagtcac 660  
 OY 721 gatggcagagaagagataaactgttgaacctgttgcagacagagagagagagagag 720  
 Db 721 gatggcagagaagagataaactgttgaacctgttgcagacagagagagagagagag 720  
 OY 781 gagcatalaaccac 780  
 Db 781 gagcatalaaccac 780  
 OY 841 ggaaggccctcgggaatcactatgtatgtccttcaagtgcgcagagcggagagaacac 840  
 Db 841 ggaaggccctcgggaatcactatgtatgtccttcaagtgcgcagagcggagagaacac 840  
 OY 901 ttattagtaaaacagattgagagaagtggttaagcttgaaacttgaaacttcttctgag 900  
 Db 901 ttattagtaaaacagattgagagaagtggttaagcttgaaacttgaaacttcttctgag 900  
 OY 961 aatgattgcattgtcagagattaaatgagagacacacacacacacacacacacacac 960  
 Db 961 aatgattgcattgtcagagattaaatgagagacacacacacacacacacacacacac 960  
 OY 1021 caacatatttctgcac 1020  
 Db 1021 caacatatttctgcac 1020  
 OY 1081 gaaataaagagcagatgaacacacacacacacacacacacacacacacacacacac 1080  
 Db 1081 gaaataaagagcagatgaacacacacacacacacacacacacacacacacacacac 1080  
 OY 1141 cgtttagcctcagac 1140  
 Db 1141 cgtttagcctcagac 1140  
 OY 1201 agcgttcagagagac 1200  
 Db 1201 agcgttcagagagac 1200  
 OY 1261 ctaactcctatagcagac 1260  
 Db 1261 ctaactcctatagcagac 1260

Db 1261 ctaactcctatagcagac 1260  
 OY 1321 aatgattgcattgtcagagattaaatgagagacacacacacacacacacacacacac 1320  
 Db 1321 aatgattgcattgtcagagattaaatgagagacacacacacacacacacacacacac 1320  
 OY 1381 aatatcagccttgaagaaggttacagaaggttgagattcaacatcactccagaatgta 1380  
 Db 1381 aatatcagccttgaagaaggttacagaaggttgagattcaacatcactccagaatgta 1380  
 OY 1441 acaataggtgtcagcactccacatctatgtgaaacacacacacacacacacacacacac 1440  
 Db 1441 acaataggtgtcagcactccacatctatgtgaaacacacacacacacacacacacacac 1440  
 OY 1501 cagatagcagccttaagagcagagacacacacacacacacacacacacacacacacac 1500  
 Db 1501 cagatagcagccttaagagcagagacacacacacacacacacacacacacacacacac 1500  
 OY 1561 ggaacatcccaagagagaggttgcctgttgagagaagcacaagatggaagagagag 1560  
 Db 1561 ggaacatcccaagagagaggttgcctgttgagagaagcacaagatggaagagagag 1560  
 OY 1621 agccttctgtcttccgcagagagagacacacacacacacacacacacacacacacac 1620  
 Db 1621 agccttctgtcttccgcagagagagacacacacacacacacacacacacacacacac 1620  
 OY 1681 agcagaatgcagatcccaagaagaagcagaagagagagagagagagagagagagag 1680  
 Db 1681 agcagaatgcagatcccaagaagaagcagaagagagagagagagagagagagagag 1680  
 OY 1741 gatggcagagaaggttctgacatltgaagtcacacacacacacacacacacacacacac 1740  
 Db 1741 gatggcagagaaggttctgacatltgaagtcacacacacacacacacacacacacacac 1740  
 OY 1801 ctgtgtgtcagtgtaaaaggttaacggttcaaaagagaacacacacacacacacacacac 1800  
 Db 1801 ctgtgtgtcagtgtaaaaggttcaaaagagaacacacacacacacacacacacacacac 1800  
 OY 1861 gtcagatccatataatgagagagacacacacacacacacacacacacacacacacacac 1860  
 Db 1861 gtcagatccatataatgagagagacacacacacacacacacacacacacacacacacac 1860  
 OY 1921 caactgatgcagtaaatgagagaatccctgttgggaagaacacacacacacacacacac 1920  
 Db 1921 caactgatgcagtaaatgagagaatccctgttgggaagaacacacacacacacacacac 1920  
 OY 1981 accctaagaaggtctatgtcttactgaagcacaataaagagaatgatacagcttatgtt 1980  
 Db 1981 accctaagaaggtctatgtcttactgaagcacaataaagagaatgatacagcttatgtt 1980  
 OY 2041 gcaagagaaataagcagaatgagatgagtgaaagtcacacacacacacacacacacacac 2040  
 Db 2041 gcaagagaaataagcagaatgagatgagtgaaagtcacacacacacacacacacacacac 2040  
 OY 2101 ctgoccatlgaac 2100  
 Db 2101 ctgoccatlgaac 2100  
 OY 2161 ggaattgagagaggttgaatgaatcccccagcagaaatgctgcccctagatgagatgag 2160  
 Db 2161 ggaattgagagaggttgaatgaatcccccagcagaaatgctgcccctagatgagatgag 2160  
 OY 2221 gagtgaagtaataac 2220  
 Db 2221 gagtgaagtaataac 2220  
 OY 2281 atagaagaatgagaggttgcagtgcttccacatctctcagacagtcacacacacacacac 2280  
 Db 2281 atagaagaatgagaggttgcagtgcttccacatctctcagacagtcacacacacacacac 2280  
 OY 2341 tcccatgatgt 2340  
 Db 2341 tcccatgatgt 2340

Db 2332 TCCATGATGATGTTGGGTTGTGACGGCAGATGCTGTAATTTGGGCCAAGGCTGCAATC 2391  
QY 2401 agtaltcaagcagcagctctcttgagtcagagatgtgataccagttctctgtctttcaacga 2460  
Db 2392 AGTGATTCAGCGCAGCTGCTTTGTAGTCCAGATGTGATCCAGTTCTTGCTTTTCAAGA 2451  
QY 2461 gaagagatttgacgtcagagatgtcagaanaaagcacaaagcaatttcagatgacagt 2520  
Db 2452 GAAGGATTTGGACGTCAGATGATGTCAAAAAACCCAAAGCAATTTTCAGATGCCAGT 2511  
QY 2521 caattggatttcgttaaacacgaanaaatacaaaagcatgattagatgattgagtcagag 2580  
Db 2512 CAATTTGGATTTGTTAAACACGAAATCAAAACCATGATGATTTAGTTAGCTTACGAG 2571  
QY 2581 actaaacccaatacagtagtagacgaagaagaagttctccacagagatgtggtctc 2640  
Db 2572 ACTAAACCATATACAGTGTGATGACCAAGAACAGGTTCTCCAGCAGAGATGTGGGTCT 2631  
QY 2641 tccctgggtctgaagaagtcgaagctcgttggaagctctgcaagccgagcttgccagtg 2700  
Db 2632 TCCCTGGGTCTGAAGAAATCAAGCTGTTGGAGAGTCTGACAGCCGCGAGTGGCCAGGTG 2691  
QY 2701 actttgaatggggaatattccctttccatctgccaagccgagatataatgaagcgagggga 2760  
Db 2692 ACTTTGAATGGGGAATATTCCTTTCCATGTCACGCGCGGATATATGAGAGCGGGA 2751  
QY 2761 tgcgaatgagacttcaagagctgcacacgaacatcttaagtaaaacccgcgtaagtat 2820  
Db 2752 TGCATATGAGACTTCAGAGCTGCCATGCAAAATCTTATGATTAACCCGCGGTAGATGAT 2811  
QY 2821 gatgaatgaagcattgagagaccttggagaagaacacagaagaagtccaagtcaaggagga 2880  
Db 2812 GATGATGAAGGACATGAGAACCTTGGAGAAACACAGAAAGAAATTTCAAGATCAGGAGGA 2871  
QY 2881 gagctgtatccaaagccagagatcagcttccacactctcggagaagcaaatgatatga 2940  
Db 2872 GAGTGTGATCCACAGCCAGATGATGAGCTTCCACTCTGTGAGACGCAAAATGATGA 2931  
QY 2941 aaccaagaagaagtgataaagacttgataaanaaagaataaactcggaanaaagaagaag 3000  
Db 2932 AACCAAGAAAGGTGATTAACATGATGAAAAAAGATTAACCTGGAATAAAGAAAGAG 2991  
QY 3001 aaagatagagataaaggagaagataaanaatgaagccaaagggaatgctgaaggcttg 3060  
Db 2992 AAAGATTAAGATTAAGAAAGATTAATGAAGCCAAAGAGGGAATGCTGAAGGGCTTG 3051  
QY 3061 ggaagacatgctcaggttggaacaaatcgaaaagaatgacaagaattgaaaacacggttaa 3120  
Db 3052 GGAGCATGTTTCAGGTTGGCAAAACATCGAAAGATGACAAAGATTGACAAAAACGGGTAAA 3111  
QY 3121 aataaatacaggaatcccttaccatcagaagaagagagatacgaatgaagcagaagcag 3180  
Db 3112 ATTAATAATACAGGAATCCTTTACATCAGAAAGAGAGAGATGAGATGACAGAGAGCAG 3171  
QY 3181 gagaggaatcaagccaaactcgagaatttgaggagacgaacagctcgagagcgtgaactat 3240  
Db 3172 GAGAGGATTCAGACCAAAACTCGAGAAATTTAGGGAAACAGCAAGCTCGAGAGCTGACTAT 3231  
QY 3241 gctaaatcaagaatttcacatcgagacatttgctgtgtagtgaattgaattgaattgggga 3300  
Db 3232 GCTGAATTCAGAGATTTTCATGCGACATTTGGCTGTGATGAGATTAATGATGGGGA 3291  
QY 3301 gttcttcttatagaagttcactgagctcacaagctagaacccaagaagcccaagaagaag 3360  
Db 3292 GTTCTCTTATGAAGGTTCATGGCTCTCAACGCTAGACCTCAAGGCCACAGAAAGGG 3351  
QY 3361 catatgagagatgcttctgtatgcccgaagtcaagaagccgaggaattccaacccctcact 3420  
Db 3352 CATATGATGATGCTTTTATGCCCCAAAGTCAAGAGCCGCGGAATTCAAAACCTCACCT 3411  
QY 3421 gtagacagtaacagatacaactccctagcaatcatgatacgatacaacgctctgagcaagaa 3480  
Db 3412 GTAGACATTAACAGATCAACTCTTAGCAATCATGATCGGATACAGCGTCTGAGGCCAAGAA 3471

QY 3481 ttccagcaagcaaaagcaagatgataagaatgataagaatcgtcggcggaacctatagtttag 3540  
Db 3472 TTTGAGCAAGCAAAACAGATGATGATGATAAGATCGTGGCGGACCTATGTTTGGAG 3531  
QY 3541 caaaccttgccgaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 3600  
Db 3532 CAACCTTGCCGACACGACGCGCGCGCGACGACGCGCGCGCGACCTCGGTGTCCTGGAG 3591  
QY 3601 gtcagatgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 3660  
Db 3592 GTGCAATGACAGCGGACGCGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 3651  
QY 3661 tacagctctctgctcctcggcaaaagcaagaanaatgcaagctcgtctcccaagactcttg 3720  
Db 3652 TACAGCTCTCTGCTCGGCAAGAGGAAATGCAAGTCCGCTCCAGAGACTCTTG 3711  
QY 3721 gacgaagactactcccttgagggaagcttcaagatgccaagaagaacccaagtaactcc 3780  
Db 3712 GAGCAGAACTACTCCCTCGGAGAGGCTTCCAGAGTGCCTAAAGAGAACCCAGTACTCC 3771  
QY 3781 agctacaaagcttccaggaagagctactcgtgagagacatgcttcaagccagaggtcatg 3840  
Db 3772 AGCTACCAAGGCTCCAGAGAGGCTTACTGGAAGGACATGCTTCAACCCAGGTCATG 3831  
QY 3841 ctggaactcgaagagctccttcgccaagagacagagcggaagaagcagacagatgaagaag 3900  
Db 3832 CTGGAAGCTCAGAGACTCTCTTCCAGAGAAAGAGGCGGAAGAGAGAGATGAAGAAG 3891  
QY 3901 cagctccttcgaggggccgaacataatgactgataagaagaattccagaacccagct 3960  
Db 3892 CAGCTCTCTTCCGAGGGGCGCCAGCAACTATGACTGTATTAAGAAAGTCCAGAGCCCACT 3951  
QY 3961 tacgccccctccaaaggagcccttcggcaagaatgagccccctccctcctaagtttgg 4020  
Db 3952 TAGGCCCTCCCAAGGGGCTTCCGCGAGATGTCCTCCCTCTCTCAGGTTGG 4011  
QY 4021 aggcctgaacagactcagactccttgagaagaagagccctctatctctga 4071  
Db 4012 AGGCTGAACAGACTTCAGACTCTGAGAAAGGAGGAGCCCTTCTATTCTCGA 4062

RESULT 4  
AF467004 3960 bp mRNA linear PRI 07-FEB-2002  
LOCUS Homo sapiens partitioning-defective 3 protein splice variant d  
DEFINITION (PAR3) mRNA, complete cds, alternatively spliced.  
ACCESSION AF467004  
VERSION AF467004.1 GI:18568347  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 3960)  
AUTHORS Gao,L., Macara,I.G. and Joberty,G.  
TITLE Multiple splice variants of PAR3 and of a novel related gene,  
PAR3L, produce functionally different proteins  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3960)  
AUTHORS Gao,L., Macara,I.G. and Joberty,G.  
TITLE Direct Submission  
JOURNAL Submitted (10-JAN-2002) Center for Cell Signaling, University of  
Virginia, PO Box 800577, Charlottesville, VA 22908, USA  
FEATURES  
source location/Qualifiers  
1..3960 /Organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..3960 /gene="PAR3"  
/note="PAR3"  
1..3960 /gene="PAR3"

```
/note="PAR3D: multi-PDZ domain containing protein;  
alternatively spliced"  
/codon_start=1  
/product="partitioning-defective 3 protein splice variant"  
/protein_id="AAL7604.1"  
/db_xref="GI:18568348"  
/translation="MKVYCGFRTVVVPCGDGHNKVFSLIQAVTRYKATKADNY  
WIOVHLEHEDGGIIDLDDICDVAADDRLVAFDEDDPHGGDGTGASTOSTPE  
IFSELDLNNVSAFOPYONSETIETVPSVLAVFDEDDPHGGDGTGASTOSTPE  
RSSEPSKPNKTRKSTAGTFLKONTAGSPKTCDRKDKENDYSLRDSNSNDQSDNA  
SSLSASHMVKWLEKRODDEDEEDNSVEVEKVGADHGLPHINESLDMDKIVE  
VPNDGPGILGIHVPPSSAKGSGTGLVLEKKGKAEHGLPHINESLDMDKIVE  
RPEOAHMFROAMRPLIMEHVVPANKROYDLSQKNNNTSSREPSDQYIDNR  
YNNKTKGLNLTOLKKEGEGISITRDVTIGSPITVKNILPBGALIDGRLKAG  
KETTKEGNDLVLPDGTREFLEPEVPLNDGSGSLVFKOEDAFHRENNASPSOMD  
INGKASDQGLRYNDQLIANGESILGKTDNDAKMTLKRSMSTGKRCMAIIVAK  
RISKCNELKSPGSPGDELPIETALIDRERKISHLSYSGIEGLDESPRNAAISRIK  
ESGKYOISPTVNMFPDDTVIIEEDRLPRLPHLSQSSSSSHDDPYVTADAGTWA  
AISDSADSLSPVDYPIAQRGREGROKMSERTKOPSDASOLDPYTKRKSMIDG  
IADETKINTYDDQKAGSPSDVPSGLKSSLSLEIQTAVAETPLNGDIPFHRPFR  
IIRGCGNESFRAIDKSTDKPAVDDDEGEMTLEEDTESSRSRGSSTASDQSH  
SIEROMGNOEKEDKTDKDKRTGKKEKDKDKDKDKDKDKDKDKDKDKDKDKDK  
REPEROREREDAEIODEHRTGCDDELATGSGSYEGSALMLKGLDMPRIQKT  
LYAOVKPRKRSFSPVDSNRSTPSNDRIOFLROEPQAKDDVEDDRRTYSFEDPW  
PNARPAQSGRHSVEVOMOROEERESSOOKORQISLSPDSHRKNASVSODSWE  
ONYSPGEFQSAKENPRYSYSGSRNGTGHGNAVMTLEIOLROKROKBOAK  
KOPSPGSPSNYSYKRVDPSPVAPKPGFPODVPSPSOVARLRLQTPKRRPYS"  
814..1074  
/gene="PAR3"  
/note="Region: PDZ domain"  
1378..1638  
/gene="PAR3"  
/note="Region: PDZ domain"  
1765..2028  
/gene="PAR3"  
/note="Region: PDZ domain"  
BASS COUNT 1170 a 910 c 1048 g 832 t  
ORIGIN
```

Query Match Best Local Similarity 53.2%; Score 3024; DB 9; Length 3960;  
Matches 3074; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 atgaagtgaccgtgcttcggaagaccggtgctgctgcgctgcggaagccac 60  
DB 1 ATGAAGTGACCGTGTGCGACGGACCCGGTGTGCGTGTGCGGCGGACGCCAC 60  
QY 61 atgaagtttcagcctcaccagagcggtgacccgtaccggaagccatcgccaag 120  
DB 61 ATGAAGTTTTCAGCCCTCATCCAGCAGCGCGGTGACCCGCTACCGGAAGCCATCCGCCAAG 120  
QY 121 gatccaaactactgatacagtgatcgcttggaacatgagatggaagatataaac 180  
DB 121 GATCCAAACTACTGATACAGTGCATCGCTTGGAACATGAGAGATGAGAAATCTATAC 180  
QY 181 ctgtatgacattcttctgtatgtagcagaagataaagaagactgtgacagtttgat 240  
DB 181 CTGTATGACATTCTTCTGTATGTAGTAGCAGACGATAAAGACAGACTGTAGCAGTGTGAT 240  
QY 241 gagcagagatccaatcaggaagtgatgacacagtgccagttccaaggttaccagagc 300  
DB 241 GAGCAGAGATCCCAATCAGGAGGTGATGACACCACTGCCAAGTTCACCGGTTACCCAGAGC 300  
QY 301 ccagagatatctgtatgtagcttggaaccaaatgtctcaacgtttacgcttaacca 360  
DB 301 CCAGAGATATTGTGATGAGCTTGACCAACATGTCTCAACGCTTTAGCCTTACCAAA 360  
QY 361 gcaacagtggaattgaggtcaccacttccttcgagcaaatatgctcttaattt 420  
DB 361 GCAACAGTGGAATTGAGGTCAACCTTCCTTCGAGCAAAATATGCTCTTACATTT 420
```

---

```
DB 361 GCAACAGTGGAATTGAGGTCAACCTTCAGTCTTTCGAGCAAAATATGCTCTTACATTT 420  
QY 421 cgacagtagtagccagactcaattgagcctcctcactctcgtcagtagtagtaattt 480  
DB 421 CGACAGTAGTAGCCAGACTCAATTGAGCCTCTCCTCCTCTGTCAGTAGTAGTAATTTT 480  
QY 481 tctcttgaagagcttcaaggaataatcccaacagcttggtcaacaacagctgtcttc 540  
DB 481 TCTCTGAAAGAGCTTCAAGGAATAATCCCAACAGCTGTGTCACACAGCTGTCTTCTTC 540  
QY 541 aagcagaacacgtcgtggaggtctcttaaaccccgacagcaggaagaagaatgaataacaga 600  
DB 541 AAGCAGAACACGTCGTGGAGGTCTCTTAACCCGACAGCAGGAAGAAGAATGAATAACAGAGA 600  
QY 601 agcctccgcggagatagtagtaactggtctcaacaaattccagagagacaatgctgcctc 660  
DB 601 AGCCTCCGCGGAGTAGTAGTAAGTGTCTCAACAAATTCACAGAGACAAATGCTCGCTCG 660  
QY 661 tctctgagtgccagtcacccaatggtggtgcaagtgcttgagagaacacagatgag 720  
DB 661 TCTCTGAGTGCAGTCAACCCAAATGTTGATATGTTAAAGCTCTGTAAGATCCCAACAAAT 720  
QY 721 gatggagcagaagagatatacagctcgttgtagaactgttgtagactgtcgaacaggttg 780  
DB 721 GATGGAGCAGAAAGAGATATACAGTGTGTGAACTGTTGAACTGTTGAACTGTTGAACTGTTG 780  
QY 781 gagcatataccaacttctcttgatgataatgtaagctcgtlaagaagtcaccaagat 840  
DB 781 GAGCATATACCACTTCTCTGATGATATGTAAGCTGTTGAACTGTTGAACTGTTGAACTGTTG 840  
QY 841 gtagggcctctggaatcatgtagtgcttcaagtgctcgaagcggaggaacccctggg 900  
DB 841 GTAGGGCCTCTGGAATCATGTAGTGCTTCAAGTGTGCGAGGGCGGAGAACCCCTGGG 900  
QY 901 ttattagtaaacgatttggaagaagtggtgaagctggaacatgaaatcttttcgtgag 960  
DB 901 TTATTAGTAACGATTGGAAGAAGTGGTGAAGCTGGAACATGAAATCTTTTCTGTGAG 960  
QY 961 aatgattgcatgtgcaggaatgataatgtagcagcttcgaataagaagatttgaagaaga 1020  
DB 961 AATGATTGCATGTGCAAGATTAATGATGAGCAGCTTCGAATAGAGATTGGAACAGCA 1020  
QY 1021 caacatagtttcgcaaacctgctgtaacaccatcatcttggttcattggttctctga 1080  
DB 1021 CAACATAGTTTTCGCAAACCTGCTGTAACCCATGCTATGTTGTTCTCATGAGTCTCTCA 1080  
QY 1081 gcaaatgaagacagtatgaacaactaccacaagtggaagaagaacatctactatcaagc 1140  
DB 1081 GCAAAATGAAGACGATATGAAACAACCTATGCCAAAGTGGAAGAACAATTAATTAATCAAGC 1140  
QY 1141 cgttttagccttgacagccagatattgcaacaagaagtggtgaagcgtgagaggttcac 1200  
DB 1141 CGTTTTAGCCTTGACAGCCAGATATTGCAACAAGAGTGTGAACAGTGCAGGAGCTTCAC 1200  
QY 1201 acggtgcaagagaaccccgacttgaacccacccgctggaagatagactctcaataaga 1260  
DB 1201 ACGGTGCAGAGAACCCCGACTTGAACCCACCCGCTGAGAGATAGACTCTCACTCAAGA 1260  
QY 1261 ctactctatagcgacacccctcgggaacacacacacacacacacacacacacacacacac 1320  
DB 1261 CTACTCTATAGCGACACCCCTCGGGAACACACACACACACACACACACACACACACACAC 1320  
QY 1321 aatgatttagtagacactgtagaagtggttataacacacacacacacacacacacacacac 1380  
DB 1321 AATGATTTAGTAGACACTGTAGAAGTGTGTTATTAACACACACACACACACACACACAC 1380  
QY 1381 aatatcagcttaagaaaggttacagaaggtttggaattagacatcaactccaagatgta 1440  
DB 1381 AATATCCAGCTTAAGAAAGGTACAGAAGGTTTGGATTGACATCACTTCCAGAGATGTA 1440  
QY 1441 acaatagtgtagcagctccaatctatgtgaacaaacatctcccccggggggggggccatt 1500  
DB 1441 ACAATAGTGTGCTCAGCTCAATCTATGTGAAACAAATTCCTCCCGGGGGGGGGCCATT 1500
```

Oy 1501 cagatgagcgaactaaaggcagagacagacttaagaagtaaatgagagtagattagtg 1560  
 |||||  
 Db 1501 CAGATGCGCCACTTAAGCGAGAGACAGACTTATAGAGGTAAATGAGATGATTAGTG 1560  
 Oy 1561 ggcgaatcccaaggaggaagttgcttgcgtgtttgagaacacccaagaatggaagaaactgtg 1620  
 |||||  
 Db 1561 ggcgaatcccaaggaggaagttgcttgcgtgtttgagaacacccaagaatggaagaaactgtg 1620  
 Oy 1621 agccttcgtcttcctgcgcagagaagccttcacccaaggagaactgaatgcagagcca 1680  
 |||||  
 Db 1621 AGCCTTCGTGCTTCTTGCGCAGAGAAAGCCCTTCCACCAGGAAGACTGATTCAGAGCCA 1680  
 Oy 1681 agccagatgcagatcccaaaagaaacgaagcagaagaatggaagatactgttccacact 1740  
 |||||  
 Db 1681 AGCCAGATGCAGATTCCAAAGAAACGAAGCAGAAATGAGATATTGTTCTACACT 1740  
 Oy 1741 gatggcaccagaggaaattctcgaacttgaagtcacacttaagtatcagagtcagagc 1800  
 |||||  
 Db 1741 GATGGCACCGAGGAATTTCTGCATTGAGTCCACTTAATGATTCAGGATCTGCAGGC 1800  
 Oy 1801 ctgtgtcagtgatcaaaaggttaacccggtcaaaagagaacccagcagatttggaactctt 1860  
 |||||  
 Db 1801 CTGTGTCAGTGTCTAAAGGTAAAGGTACCGGTCTAAAGAGAACACGCGATTTGGAAATCTTT 1860  
 Oy 1861 gtcaagtcacatattatgaagaggaagcagacatctaaagaatggaagcctcgagtgagat 1920  
 |||||  
 Db 1861 GTCAAGTCCATTATTAAAGGAGAGACACATCTAAAGATGGAAGGCTTCGGGTGAATGAT 1920  
 Oy 1921 caactgataagcaatgaatggaggaatccctgtttggcagaagacaacaaagatgcattgaa 1980  
 |||||  
 Db 1921 CAACGATAGCAAGTAAGAAAGGAATCCCTGTGGGCAAGACAAACAGATGCTCATGGA 1980  
 Oy 1981 accctaagaaggtcattgtctactgaaggcaataaagcagaagaatgcacacttaattgt 2040  
 |||||  
 Db 1981 ACCCTAAAGAGGTATGTCTACTAAGGCAATAAACGAGGAATATCAGCTTATTGT 2040  
 Oy 2041 gcaagagaataaagaagtgcaatgaagtgcaatgaagtgcaatgaagtgcaatgaagtgcaatgaag 2100  
 |||||  
 Db 2041 GCAAGGAAATAAAGCAAGTGCATGCAATGAGCTGAAGTCACTGGAGGCCCTTGACACTGAG 2100  
 Oy 2101 ctgcccattgaaacagcgttgatgataagaagaagaatctccactccctcactagat 2160  
 |||||  
 Db 2101 CTGCCATTGAAACAGCCTTGATGATGAGAAACGAAGAAATTCCTCATTCCTCTACACT 2160  
 Oy 2161 ggaattgaggggctgtgaatcgccagcagaatgctgcccactgaagataatgagtg 2220  
 |||||  
 Db 2161 GGGATGAGGGGCTTGATGAAATCCCCACAGCAAAATGCTGCCCTCAGTAGAATATGGGT 2220  
 Oy 2221 gagtcaagtaaatcagcagctgtccctcactgaatgaatgccccaaagatgacactgcat 2280  
 |||||  
 Db 2221 GAGTCAAGTAAATACACACTGTCCCTCAGAGATATGCCCCAAGATGACACTGTCAAT 2280  
 Oy 2281 atagaagaatgcaaggttccagtgcttccctcactcactcctcagcagatgccctccagc 2340  
 |||||  
 Db 2281 ATAGAAGATGACAGGTCCAGTGTCTCTCCACATCTCTGACACAGTCCCTCTTCCACG 2340  
 Oy 2341 tcccatgatagtgtgaggtttgtgacagcagatgctgtgacttgagcagaagctgaatc 2400  
 |||||  
 Db 2341 TCCCATGATGATGTGGGTTTGTGACGGCAGATGCTGTGACTTGGCCCAAGGCTGCATTC 2400  
 Oy 2401 agtgaatcagcagactgctccttgagccagatgctgacagtgatccactgtgctttcaaga 2460  
 |||||  
 Db 2401 AGTGAATCAGCCGACTGTCTTGGAGTCCAGATGTGATCCAGTCTTCTTAAACGA 2460  
 Oy 2461 gaagaattgagcctcagagatgctcagaanaagcacaanaagcaatttcaatccagt 2520  
 |||||  
 Db 2461 GAAGATTTGACGCTCAAGATGTGTCAGAAACGCAAAACCAATTTTCAATGCTCAAT 2520  
 Oy 2521 caattgattctgcttaaacacgaataatcaaaagcattgattagatagctacag 2580  
 |||||  
 Db 2521 CAATTGATTTGCTTAAACGCAAAATCAAAACATGATGATTAGGATACCTACAGG 2580

Oy 2581 actaaactcaatagctggaatgacgaagaagcaggtctccacagcagatgtggtcct 2640  
 |||||  
 Db 2581 ACTAAACTCAATAGCTGATGATGACGAAGAAGCAGGTTCTCCACGACGAGATGTGGCTCCT 2640  
 Oy 2641 tccctgggtctaaagaatgcaagctcgttggagatctcgaacccgaattccgaggg 2700  
 |||||  
 Db 2641 TCCCTGGGTCTAAAGAAAGTCAAGCTCTGTGGAGATCTCAGACGCAATGCTCCGAGGTG 2700  
 Oy 2701 acttgaatgagatacttccttcacatgctccagcgcccgagataatcagagcagagga 2760  
 |||||  
 Db 2701 ACTTTGAATGGGATATTCTCTTCCATCTCTCCACGCCCGCGATATATCAGAGCGGGA 2760  
 Oy 2761 tgcgaatgagactcagagctgcacatgcacaaacttaataaaccgcggtagatgat 2820  
 |||||  
 Db 2761 TGCATGAGAGCTTCAGACTGCTCCATCGACAATCTTATGATTAACCCCGGTAGATGAT 2820  
 Oy 2821 gatgatgaagcagatgagaccttggagaagaagcacagaagaagttcaagatcagggaga 2880  
 |||||  
 Db 2821 GATGATGAAGGATGAGACTTGGAAAGACACAGAAAGAAAGTTCAAGATCAGGGAGA 2880  
 Oy 2881 gagctgtatccacagcagctgacagctccactcctcgtgagagacaatgaatgga 2940  
 |||||  
 Db 2881 GAGCTGTATCCACAGCCAGTATGATGACCTTCCACTCTCTGAGAGACAAATGATGGA 2940  
 Oy 2941 aaccaagagaaggtgataagaactgataaanaagataaactggaanaagaaagaaag 3000  
 |||||  
 Db 2941 AACCAAGAGAAAGGTGATGATGACTGATGAAAAAAGCATTAACCTGCAAGACGAAAGAG 3000  
 Oy 3001 aaagatagagataagaagaagaataaataagaagccaagaaggaatgctgagggcttg 3060  
 |||||  
 Db 3001 AAAGATGAGATGAAGAGAAAGATTAATGAAGCCMAAGAGGAATGCTGAAGGGCTTG 3060  
 Oy 3061 ggaagacatgttcag 3075  
 |||||  
 Db 3061 GGAGACATGCTCAGG 3075  
 RESULT 5  
 AK000761  
 LOCUS Homo sapiens cDNA FLJ20754 fls, clone HEP02246.  
 DEFINITION AK000761  
 ACCESSION AK000761.1 GI:7021050  
 VERSION AK000761.1  
 KEYWORDS oligo capping; fls (full insert sequence).  
 SOURCE Homo sapiens hepatoma cell\_line:Hepg2 cDNA to mRNA, clone\_11b:HEP clone:HEP02246.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (sites)  
 Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
 NEDO human cDNA sequencing project  
 Unpublished (2000)  
 2 (bases 1 to 2696)  
 TITLE  
 JOURNAL NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Depart of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).  
 COMMENT  
 FEATURES  
 source location/Qualifiers  
 1..2696







Db 468 CTTGATGACATTCCTTGTGATGTAGCAGACGATTAAGACAGACTGGTAGCAGTGTGGAT 527  
QY 241 gagcagagatccacatcaacagagatgatgacacagatgacagatgacagatgacagagc 527  
Db 528 GAGCAGAGATCCACATTCACAGAGAGATGAGCAGCAGTGCAGTTCACAGGATCCAGAGC 300  
QY 301 ccagagagatgtagttagttagttagttagttagttagttagttagttagttagttagttagt 300  
Db 588 CCAAGAGATATTTGGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 587  
QY 361 gcaacagatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg 360  
Db 648 GCACACAGTGAATTTGAGGTGACACCTTCACTTCACTTCACTTCACTTCACTTCACTTCA 420  
QY 421 ccagcagatgagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 420  
Db 708 CCAAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 707  
QY 481 tccctgaaagagccttaagaaagaaatcccaagcagcctgtaacaaagcagcctgtaacaa 480  
Db 768 TCCCTGAGAGCCTTCAAGGAAATCCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 767  
QY 541 aagacagacacgctgtaggagctcctaaacgtagcagagagagagagagagagagagag 827  
Db 828 AAGCAGACAGCTGCTGGAGTCTTAAACCTGCGACAGAGAGAGAGAGAGAGAGAGAGAG 600  
QY 601 agcctccagagagatagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 887  
Db 888 AGCCTCCCGGGGATGCTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 947  
QY 661 tctctgagtagcagtagcagtagcagtagcagtagcagtagcagtagcagtagcagtagc 947  
Db 948 TCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 720  
QY 721 gatggag 1007  
Db 1008 GATGGAG 780  
QY 781 gagagatataccacatct 1067  
Db 1068 GAGGATATACCCAACTTTCTCTGTGATGATGATGATGATGATGATGATGATGATGAT 840  
QY 841 gaggggagcctgtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1127  
Db 1128 GGAGGGGCTGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
QY 901 ttattagtaaaagcattgagaaagatgagaaagcagtagtagtagtagtagtagtagtag 1187  
Db 1188 TTATTAGTAAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 961 aatgattgcatctgtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1247  
Db 1248 AATGATTGCAATTTGAGATTAATGATGATGATGATGATGATGATGATGATGATGATG 1020  
QY 1021 caacatatagttgcgaagcagtagtagtagtagtagtagtagtagtagtagtagtagtag 1307  
Db 1308 CAACATATGTTTGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1080  
QY 1081 gcaaaataaagcagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1367  
Db 1368 GCAAAATAAGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
QY 1141 gctttagcctgtagcagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1427  
Db 1428 GCTTTAGCCTGTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1200  
QY 1201 acggtgag 1487  
Db 1488 ACGGTGAG 1260  
QY 1261 ctaccatataagcagacacccctcgagaaacacacacacacacacacacacacacacac 1547  
Db 1548 CTACCTATAGGACACCCCTCGGGAACACACACACACACACACACACACACACACACAC 1607

QY 1321 aatgattagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1380  
Db 1608 AATGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1667  
QY 1381 aatatacagcagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1440  
Db 1668 AATATCCAGCTTGAAG 1727  
QY 1441 acaatagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1500  
Db 1728 ACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
QY 1501 cagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1787  
Db 1788 CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1847  
QY 1561 ggcacatcccaag 1620  
Db 1848 GGCACATCCCAAG 1907  
QY 1621 agcctctgagcctctgtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1680  
Db 1908 AGCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1967  
QY 1681 agcagagtagcagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1740  
Db 1968 AGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2027  
QY 1741 gatggag 1800  
Db 2028 GATGGAG 2087  
QY 1801 ctggtgtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1860  
Db 2088 CTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2147  
QY 1861 gtagagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1920  
Db 2148 GTCAAGTCTATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2207  
QY 1921 caactgtagcagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1980  
Db 2208 CAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2267  
QY 1981 accctlaagagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 2040  
Db 2268 ACCCTTAAAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2327  
QY 2041 ggaag 2100  
Db 2328 GCAAG 2387  
QY 2101 ctgccaatgaaacagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 2160  
Db 2388 CTGCGCATTTAAAGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 2447  
QY 2161 ggaattgagagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 2220  
Db 2448 GGAATTGAG 2304  
QY 2221 gtag 2280  
Db 2505 -----GGTAAATACAGCTGTCCCTACAGTGAATATGATGATGATGATGATGATGAT 2558  
QY 2281 atagaagtagcagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 2340  
Db 2559 ATAGAAGATGACAG 2618  
QY 2341 tcccatatagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 2400  
Db 2619 TCCCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2678



OY 2401 agtgcacgacgacgtcctcttgagtcacagatgttgacagcttgcttgcttccacga 2460  
|||||  
DB 2679 AGTGATGACGCGCGATGCTCTTGAGTCCAGATGTGATCCAGTCTCTTCCATCAACA 2738  
OY 2461 gaagatcttgagcgtcagagatgtgtcagaacaaacgacaaacaaatcttccaatgccaat 2520  
|||||  
DB 2739 GAAGGATTTGGACGTCAGGATGTGTCAGAAAAACGACAAACCAATTTTCAATGCGCACT 2798  
OY 2521 caattggatttcgttaaaacacgaacatcaaaacgatgttaagatagatagctgacgag 2580  
DB 2799 CAATTGGATTTGTTAAACACGAAATCAAAACATGATTTAGTGTACTGACGAG 2858  
OY 2581 actaaactcaatcacgtgtgatacgaagaagagcttctccacagagatgtgagctct 2640  
DB 2859 ACTAAACCTCAATACAGTGTGATGACCAAAAGCAGGTTCTCCACAGAGATGTGGGTCT 2918  
OY 2641 tccctgggtctcfaaagaafcaagctcgttgagagctgtcagaacgacgttgccgaggtg 2700  
DB 2919 TCCTGCGGTCTCAAGAAAGCAAGCTCATTTGAGAGCTGTGACAGCCGACGTTGCCGAGGTG 2978  
OY 2701 acttgaaatggagataatctcttcacatcgltccacgagcgcgagataatcagaagcagggag 2760  
DB 2979 ACTTTGAATGGGATATTCCTTCCATGTCACGCGCGGATATATCAGAGGACAGGGA 3038  
OY 2761 tgcgaatgagagcttcagagctgcacatgcacaaacttatgataaacccggttagatgat 2820  
DB 3039 TGCATGTGAGAGCTTCAGAGCTCCATCGCAATCTTTATGATTAACCCCGGTAGATGAT 3098  
OY 2821 gattgtgagagatgagagaccttgagaagaacacagaagaagttcaagatcagaaggaga 2880  
DB 3099 GATGATGAAGGATGAGAGCTTGGAAAGACACGAAAGAAATTCAGATCAGGAGAGA 3158  
OY 2881 gactgtgatccacacgacgagtgatcagaccttccactctcttgagagacaaatgaaatga 2940  
DB 3159 GAGTGTGATTCACAGCGCGATGATCAGGCTTCCACTCTCTGAGAGACAAATGATGTA 3218  
OY 2941 aaccaaagagaaagtgtgataagactgatagaacaaagataaaactggaagaagaagaag 3000  
DB 3219 AACCAAGAGAAAGGTGATGAGCTGATGAAAAAGATTAACCTGAAAAAGAAAGAGAG 3278  
OY 3001 aaagataagataaagaagaagataaataagaaagcagaagaaggaatgctgaaggcttg 3060  
DB 3279 AAAGATGAGATGAGAGAGATTAATAATGAAGCCAAAGAGGAAGTCTGAAGGGCTTG 3338  
OY 3061 ggaagacatgctcag 3074  
DB 3339 GGAGACATGTTGAG 3352

RESULT 7  
AL591464 196374 bp DNA linear HTG 19-DEC-2001  
LOCUS Homo sapiens chromosome 10 clone RP13-302D10, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 14 unordered pieces.  
ACCESSION AL591464  
VERSION AL591464.2 GI:17973978  
KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS Burton,J.  
TITLE Direct Submission  
JOURNLS Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk  
On Dec 20, 2001 this sequence version replaced gi:14141513.  
COMMENT  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk

----- Project Information  
Center project name: B302D10  
----- Summary Statistics  
Sequencing vector: plasmid; L08755; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 189117 bases at least Q40  
Consensus quality: 192136 bases at least Q30  
Consensus quality: 193772 bases at least Q20  
Insert size: 195074; sum-of-contigs  
Insert size: 221513; 17.0% error; agarose-fp  
Quality coverage: 4.57x in Q20 bases; sum-of-contigs Quality  
Coverage: 4.33x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 40923: contig of 40923 bp in length  
40924 41023: gap of 100 bp  
41024 52424: contig of 11401 bp in length  
52425 52524: gap of 100 bp  
52525 56668: contig of 4144 bp in length  
56669 56768: gap of 100 bp  
56769 66064: contig of 9296 bp in length  
66065 66164: gap of 100 bp  
66165 74102: contig of 7938 bp in length  
74103 74202: gap of 100 bp  
74203 118025: contig of 43823 bp in length  
118026 118125: gap of 100 bp  
118126 120523: contig of 2398 bp in length  
120524 120623: gap of 100 bp  
120624 134346: contig of 13446 bp in length  
134347 134446: gap of 100 bp  
134447 143138: contig of 8692 bp in length  
143139 143238: gap of 100 bp  
143239 151820: contig of 8582 bp in length  
151821 151920: gap of 100 bp  
151921 179327: contig of 2767 bp in length  
179328 179427: gap of 100 bp  
179428 187106: contig of 7679 bp in length  
187107 187206: gap of 100 bp  
187207 193223: contig of 6017 bp in length  
193224 193323: gap of 100 bp  
193324 196374: contig of 3051 bp in length.  
Location/Qualifiers  
1. 196374  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP13-302D10"  
/clone\_1fb="RP13-13.2"  
1. 40923  
/note="assembly:fragment:00704  
fragment\_chain:1"  
41024. 52424  
/note="assembly:fragment:02076  
fragment\_chain:1"  
52525. 56668  
/note="assembly:fragment:02444  
fragment\_chain:1"  
56769. 66064  
/note="assembly:fragment:00688  
fragment\_chain:1"  
66165. 74102  
/note="assembly:fragment:01900  
fragment\_chain:1"  
74203. 118025  
/note="assembly:fragment:02289

```

misc_feature      fragment_chain:1"
                  118126..120523
                  /note="assembly_fragment:01018
misc_feature      fragment_chain:2"
                  120624..134346
                  /note="assembly_fragment:01079
misc_feature      fragment_chain:2"
                  134447..143138
                  /note="assembly_fragment:00344
misc_feature      fragment_chain:2"
                  143239..151820
                  /note="assembly_fragment:01621
misc_feature      fragment_chain:2"
                  151921..179327
                  /note="assembly_fragment:01247
misc_feature      fragment_chain:3"
                  179428..187106
                  /note="assembly_fragment:01160
misc_feature      fragment_chain:3"
                  187207..193223
                  /note="assembly_fragment:01995
misc_feature      fragment_chain:3"
                  19324..196374
                  /note="assembly_fragment:02022
BASE COUNT      57987 a 41071 c 41893 g 54114 t 1309 others
ORIGIN
Query Match      35.2% Score 2003; DB 2; Length 196374;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2003; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3677 gcaaaagcagaaatccagctcgcgtccccaagactctggagacgaactacccc
Db 190006 GCGAAAGCAGGAAAAATGCCAGCTCGGTCTCCAGACCTCTGGACCAAACTACTCTCC
QY 3737 ctggggaaggtctccagagtgcccaagagaaccccaagctaccagctacaggtctcca
Db 190066 CTGGGGAAGGCTTCACAGATGCCAAAGAACCCCAAGTACTCAGCTACCAAGCTCCA
QY 3797 ggaacgctactctggagagacttgctcaagcagaggtgcagctggtgaagctcga
Db 190126 GGAAGGCTACCTGGGAGAGACATGCTTCAACGCCAGGGCATGCTGGAACCTCAGAGC
QY 3857 tcttcgcaagcaagagcggaagagcagagatgaagaagcagctcctccgaag
Db 190186 TCTTTCGCCAGGAACAGGCGGAGAGACAGAGATGAAGAGCAGCCCTCTCCGAGG
QY 3917 ggcacgaactatgactcgtataaagaagtcagaccaggtlaagccctcccaag
Db 190246 GGCCCAAGCACTATGACTGCTATTAAGAAAGTCCAGAACCCCAAGTACGCCCTCCCAAGG
QY 3977 ggccttcggcgaagatgtgcccctccctcctcctcaggttcgaggtcgaacagacttc
Db 190306 GGCCCTTCGGCAAGATGCCCCCTCCCTCTCTCAGGTTCGAGGCTGAACAGACTTC
QY 4037 agactctgaagaagagagagccttcctatcctcagcagcagaataaagagatgctcatg
Db 190366 AGACTCTGAGAAAGGAGGCCCTTCCTATTCCTGAGACAGCAATTAACGATGCTTCAG
QY 4097 tcggcacaataaagaacttctcctatgaagaacttgatttgggaagttttttaaaccct
Db 190426 TCGCGCATATAAAGACATTTCTCTATGAAGACTGTATTATTTGGAGATTTTAAACCT
QY 4157 cgatggtactatgagatcttctgttctgtatcagtcgcttaaggggttaaggaaa
Db 190486 CGATGGTACTATAAGAGATTTCTGTGTGTATGATGCTCTTAAAGCGGTGAAGCAA
QY 4217 gaatatgaagccttaagtcttctgacactatgtctcaagtgtctgttcaatgaagaat
Db 190546 GAAATGGAAGGCTTAATGCTTTGCCACTATGCTCAAGGTGCTCTGTTTCATGAAGAGAT

```

```

QY 4277 ttccacccctgtgaacatcatctgtttgaggtcttataatgtctcgcctcccaagt
Db 190606 TTCCACCCCTGTGACAAATCATCTGTTTGAAGTGTATATGCTCGGCCCTTCACAGT
QY 4337 accagaatctcggccctacatcatgattgtcgcggcttggtgtatcatctccagcc
Db 190666 ACCAGAAATCTCGCCCTACTCATGAGTGTCCGGCGCTTGTTGAATCCCTGCAAGT
QY 4397 acttcagtgacaatctcaacctggaagtggagagatgacgtcggccctgttccctcta
Db 190726 ACTTCAGTGACAAATTCACCTGAAGTGAAGATGACGTCCGGCCCTGTTTCTCCTCTA
QY 4457 agtctcttaagctatggagatgacatcttaagctcgtgtgagaaagtgggagacatc
Db 190786 AGTCTCTTACGTATGGATGACATTTAGTCTCTCTGTGAGAAAGTGGGACATAC
QY 4517 accaaaattggggcttctcgtactcaagaagaagccatttctgacttgcataca
Db 190846 ACCAAAATGCGGCTTCTGTACTTCAACAGCACAGCCATTTGCTACTTGTATCA
QY 4577 ctgtggttctccttctcctcctcctcctcctcctcctcctcctcctcctcctcctc
Db 190906 CTGTGTTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCT
QY 4637 agaaagctaaagcagatgcacacatgttctgacagagaggtgtcactatctac
Db 190966 AGAAGCTAAGCCATGATGTCACACATTTGTTGTGTCACCATGGGGCTTTTGTACTCTCCT
QY 4697 tatgaactaaacttcaacaaataaagtatgactcgtcgtcgttatacactagctgtc
Db 191026 TATGACTCAATTAACCAATTAAGTATGATCTCTGCTGTGGTTATATACCTCTGTT
QY 4757 cgtctcttctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc
Db 191086 CGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
QY 4817 cagagatgtgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc
Db 191146 CAGAGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
QY 4877 ggaagactaaagaagaattctcctcctcctcctcctcctcctcctcctcctcctcctc
Db 191206 GGAGAGCTTAAGAAATATCTCTAGGAAGCTCATGSCCAGTCAATCTATGTAAT
QY 4937 atttaattagtttttctcctcctcctcctcctcctcctcctcctcctcctcctcctc
Db 191266 ATTATTAATGATTTTGTCTGACAGCTTTCAGGAAGGCGACAGAAATGGACAGATA
QY 4997 aacagacagtcatttctcctcctcctcctcctcctcctcctcctcctcctcctcctc
Db 191326 AACAGACATGATATTTTATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
QY 5057 tgcctaatgcatccacacagatgtctcgaagcggatgacagcctcctcctcctcctcctc
Db 191386 TGTCTAATGATCCACACACTGATGTCTGACCTGGCATGCCAGCTCTAACTAGTCT
QY 5117 gccatgaacaactaaagaagaagaagccgttacaatcaacacagagaagaacttttcg
Db 191446 GCATGAACAATAAAGACAAAGAAAGCCGTTTACACATCACACAGAACATTTTGG
QY 5177 gtcccaagagtggtgtggaagaagcactcctcgtcagtaattgagtggtgtgtgtg
Db 191506 GTCCCAAGAGGTGGTGGAGAGATCTCTCGGTCAGATTAGAGTGTGTGTGTG
QY 5237 gttctggagatctcgttgcctccatcttctcctcctcctcctcctcctcctcctcctc
Db 191566 GTCTCGGAGATCTCGGTGGCTCCCATCTCTCTTCAATGTTCTGAAACATCTGTATGTGA
QY 5297 accatgctcggggtcgaagtgctgtgaatcccgatgtggaagaagctggaggtgtaa
Db 191626 ACCATGCTGGGCTCTAAAGTGCTCTGGAATCCCGATGTGGAAGAAAGCTGAGGTGAA

```

OY 5357 gctcagcataccatgtatttactttaaaacagaaacacatgtatgatgtctta 5416  
 DB 191686 GCTCAGCATACCATGTATTCTTTAAAAACGAAAAACATGATGTGATGTGTA 191745  
 OY 5417 ttttttttttggacattgtatttctgttgactgtcttttagaatgatgtcc 5476  
 DB 191746 TTTTtttttttttGGACATGTATTtTtGTGTGACtTGTttTAGAAATGATGTCTCC 191805  
 OY 5477 acacagtaaccgctgtctctctcgaattctcgtgtcaltgatgtcttcttaacagtg 5536  
 DB 191806 ACACAGTACCCGCTGCTCTCTGATGATTTCTGTGTGATGTGTTCTTAAATCAGCTG 191865  
 OY 5537 cggcggtgtcttaagtgtgtgtaccagtgtaacgagtgacgttgatgagacgtgtcct 5596  
 DB 191866 CGGGGtGTCTTAAGTGTGTTACAGTGTACCGCAGTGAAGACCTTGAGACAGTGGCTCT 191925  
 OY 5597 ttccacagctccctcctagctgtgagaaacagcttctcgtacatgtgaactcctaa 5656  
 DB 191926 TTTTCACAGCCTCCCTCGTAGCTGTGACAAACAGCTTTCTGTGTACATGTGCACTCCTAA 191985  
 OY 5657 taaaagcatatttcttctcgtt 5679  
 DB 191986 TAAAGCATATTCTCTCTCTGT 192008

RESULT 8  
 AL160409/c 106198 bp DNA linear PRI 18-JUL-2001  
 LOCUS Human DNA sequence from clone RP11-406D17 on chromosome 10,  
 DEFINITION complete sequence.  
 ACCESSION AL160409  
 VERSION AL160409.12 GI:14970794  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 106198)  
 AUTHORS Sycamore,N.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 COMMENT requests: clonerequests@sanger.ac.uk  
 On Jul 19, 2001 this sequence version replaced gi:14625535.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em., EMBL; Sw.,  
 SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-406D17 is from the library RP11-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/Dacpac/home.htm  
 VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone  
 RP11-406D17. It may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.

FEATURES  
 source The true left end of clone RP13-348N17 is at 104199 in this  
 sequence. The true right end of clone RP11-490D24 is at 2000 in  
 this sequence.  
 Location/Qualifiers  
 1..106198  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-406D17"  
 /clone\_11b="RP11-11.2"  
 27..160  
 /note="MIR repeat: matches 3..139 of consensus"  
 989..1039  
 /note="MIR repeat: matches 93..144 of consensus"  
 1232..1275  
 /note="22 copies 2 mer ac 77% conserved"  
 1344..1651  
 /note="Alusx repeat: matches 1..302 of consensus"  
 4229..4606  
 /note="MIR17 repeat: matches 1..403 of consensus"  
 4648..4815  
 /note="FRAM repeat: matches 4..171 of consensus"  
 5215..5516  
 /note="Alusx repeat: matches 1..300 of consensus"  
 6161..6216  
 /note="MIR repeat: matches 47..104 of consensus"  
 6217..6527  
 /note="Alusx repeat: matches 1..295 of consensus"  
 6528..6637  
 /note="MIR repeat: matches 104..205 of consensus"  
 6646..6950  
 /note="Alubd repeat: matches 1..304 of consensus"  
 8061..8354  
 /note="Alubd repeat: matches 1..289 of consensus"  
 8443..8602  
 /note="Alubd/FRAM repeat: matches 152..308 of consensus"  
 10005..10072  
 /note="L2 repeat: matches 2673..2741 of consensus"  
 10542..10849  
 /note="MER57B repeat: matches 43..386 of consensus"  
 10851..10975  
 /note="FLAM\_C repeat: matches 15..133 of consensus"  
 11451..11758  
 /note="Alubd repeat: matches 1..312 of consensus"  
 12992..14459  
 /note="PTR5 repeat: matches 504..1874 of consensus"  
 14460..14550  
 /note="MER61E repeat: matches 377..466 of consensus"  
 14551..14771  
 /note="PTR5 repeat: matches 1874..8522 of consensus"  
 14772..14923  
 /note="Alusx repeat: matches 3..154 of consensus"  
 14930..15218  
 /note="Alubd repeat: matches 1..293 of consensus"  
 15223..15519  
 /note="Alusg repeat: matches 1..299 of consensus"  
 15661..15818  
 /note="MER5A repeat: matches 34..189 of consensus"  
 15868..15919  
 /note="26 copies 2 mer aa 73% conserved"  
 16133..16443  
 /note="Alusx repeat: matches 1..312 of consensus"  
 16599..16707  
 /note="L1MK4 repeat: matches 7728..7834 of consensus"  
 16712..16787  
 /note="38 copies 2 mer ta 76% conserved"  
 16825..17092  
 /note="Alubd repeat: matches 1..292 of consensus"  
 17094..17254  
 /note="Alubd/FRAM repeat: matches 134..295 of consensus"  
 17579..17889  
 /note="Alusg repeat: matches 1..306 of consensus"  
 18807..19086

```

repeat_region /note="MTRD repeat: matches 3. .238 of consensus"
19087. .19391
repeat_region /note="Alusq repeat: matches 1. .310 of consensus"
19392. .19542
repeat_region /note="MTRD repeat: matches 238. .501 of consensus"
19608. .19904
repeat_region /note="Aluio repeat: matches 1. .295 of consensus"
20333. .20642
repeat_region /note="Alusx repeat: matches 1. .310 of consensus"
20670. .20956
repeat_region /note="Aluio repeat: matches 1. .286 of consensus"
21297. .21601
repeat_region /note="L2 repeat: matches 2414. .2696 of consensus"
21857. .22175
repeat_region /note="Aluio repeat: matches 1. .311 of consensus"
22502. .22739
repeat_region /note="Alusx repeat: matches 1. .239 of consensus"
23135. .23915
repeat_region /note="LIPALD repeat: matches 5369. .6158 of consensus"
27500. .27629
repeat_region /note="LIMM3 repeat: matches 6171. .6304 of consensus"
28661. .28984
repeat_region /note="Alusq repeat: matches 1. .312 of consensus"
30001. .30132
repeat_region /note="66 copies 2 mer aa 62% conserved"
30233. .30328
repeat_region /note="48 copies 2 mer ta 65% conserved"
30365. .30644
repeat_region /note="LIMC4 repeat: matches 6633. .6940 of consensus"
30645. .31062
repeat_region /note="MSTR repeat: matches 1. .424 of consensus"
31063. .31122
repeat_region /note="LIMC4 repeat: matches 6940. .6995 of consensus"
31133. .31547
repeat_region /note="LIMC4 repeat: matches 7167. .7597 of consensus"
31648. .32167
repeat_region /note="LIPAB repeat: matches 5612. .6131 of consensus"
32275. .32352
repeat_region /note="Aluio repeat: matches 5. .282 of consensus"
32592. .32686
repeat_region /note="LIMC4 repeat: matches 7791. .7972 of consensus"
33154. .33272
repeat_region /note="MIR repeat: matches 80. .185 of consensus"
33283. .33415
repeat_region /note="Aluio repeat: matches 14. .153 of consensus"
33416. .34701
repeat_region /note="PMS repeat: matches 172. .2212 of consensus"
34713. .34840
repeat_region /note="Aluio/FRAM repeat: matches 178. .302 of consensus"
35365. .35654
repeat_region /note="L2 repeat: matches 2622. .2710 of consensus"
36067. .36155
repeat_region /note="MER30 repeat: matches 1. .220 of consensus"
37547. .37852
repeat_region /note="Alusx repeat: matches 1. .311 of consensus"
37894. .38050
repeat_region /note="MER5A repeat: matches 9. .188 of consensus"
38447. .38744
repeat_region /note="Alusx repeat: matches 5. .302 of consensus"
39063. .39193
repeat_region /note="FLAM C repeat: matches 1. .131 of consensus"
39510. .39673
repeat_region /note="MER5A repeat: matches 4. .189 of consensus"
40932. .41102
repeat_region /note="LIPAB repeat: matches 1497. .1667 of consensus"
41103. .41726
repeat_region /note="LIPAB repeat: matches 5515. .6138 of consensus"
43386. .43747
repeat_region /note="THEB repeat: matches 1. .364 of consensus"
44459. .44771
repeat_region /note="Alusq repeat: matches 3. .306 of consensus"

```

---

```

repeat_region 44991. .45302
/note="Aluio repeat: matches 1. .304 of consensus"
repeat_region 45536. .51640
/note="LIPAB repeat: matches 11. .6145 of consensus"
repeat_region 51663. .51964
/note="Aluio repeat: matches 1. .308 of consensus"
repeat_region 51966. .52183
/note="109 copies 2 mer aa 65% conserved"

```

Query Match 34.3%; Score 1952; DB 9; Length 106198;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2002; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 3677 ggcacagcaggaagaaatgccagctcagtcctccagagctctggagagaaactactcc 3736
DB 89607 GGCNAAGCAGAAAAATGCCAGCTCGGTCCAGACTCTTGGAGCAGAACTACAGCC
QY 3737 ctggggaagcttcacagatgccaagaagaacccaggtactccagttaccaaggtccca 89548
DB 89547 CTGGGAAGGCTTCAGAGAGGCCAAGAGAACCCAGGTACTCCAGCTACCAAGGCTCCA
QY 3797 ggaacggctacctggagagacatgcttcaacgcagaggtcatgctgaaactcagagc 89488
DB 89487 GGAACGGCTACTGGAGGACATGGCTTCACGCCAGGTCATCTGAAACTCAGGAC
QY 3857 tccctgcaggaagaaagggcggaagagcagatgaaagaagcagctctccagag 89428
DB 89427 TCCTTCGCCAGACAGAGAGCGGGAAGAGCAGCAGATGAAGAAAGCCCTCTCCAGG
QY 3917 ggcacagcactatgactcgtataagaagtcacagaacccagttacgcccctccagag 89368
DB 89367 GGCACGACACTATGACTCGTATGAAGAAATCCAGAGCCCGATACGCCCTCCCAAG
QY 3977 ggccttcggcaagatgctcccccctccctcctcaaggttcgaggttcgaagacttc 89308
DB 89307 GGCCTTCGGCAAGATGCTCCCTCCCTCCCTCTCAGGTTCGAGGCTGAACGACTTC
QY 4037 agactccctgaaaggaagggcccttacttctcctgagcagcgaataagagatgctcatg 89248
DB 89247 AGACTCCTGAAAGGAGAGGCCCTTCTATCTGAGCAGCGAAATACGATCTTCATG
QY 4097 tcgcgaataaagaacattctctatgaagactgatttggagtttttaaacct 89188
DB 89187 TCGCGAATAAAGACATTTCTCTATGAAGACTGTATTTGGAGTTTAAANACT
QY 4157 cgaatgactatgagatattcctgltgltgltacagtgctttaaagcggtgaagcaaa 89128
DB 89127 CGATGTACTATGAGATATTCTGTGTGTGATACAGTGCCTTTAAAGCGGTATGACAA
QY 4217 gaaatggaagccttaatgltcttgccactatgctcaagtgctgtlcalgagaagat 89068
DB 89067 GAAATGGAAGCCTTAATGCTTGGCACATGTCTCAAGTGTCTGTTCAATGAGAGAT
QY 4277 ttcccaacctgtgacaatctctgttgagtgltatagtcgtcgtccctcccaagt 89008
DB 89007 TTCCCAACCTGTGACAAATCTGTTTGAAGGTGTCAATGTCTGCGCCCTCTCAACAGT
QY 4337 accaggaatcggccctactcaatgagtggtcgcgaggttggttaaacctccagc 88948
DB 88947 ACCAGGAATCTGCGCCCTACTCATGAGTTGTCGCGGCTGTGTGAACATCCCTGCAC
QY 4397 actgcaagtagacaattcactcgaagtgagaatgacgtgcgcgcgtttccctctca 88888
DB 88887 ACTTGCAAGTAGACAATTCACCTGGAAGTGAAGAGACGTGCGGCCCTGTTTCCCTCTCA
QY 4457 agttccttagctatggatgacatcttactgltgltggaagaaagtggcgacatcac 88828
DB 88827 AGTTCTCTTAGCTATGGATGACATCTTACTCTGTGTGAGAGAAAGTGGCGACATAC
QY 4517 accaaaaattgggcttctggtactcacagacagcattgtcgtacttgcatac 88768
DB 88767 ACCAAAAATTGGGCTTCTGTGACTTCAAGCACAGGCATTTGTGATCTTGTATCA

```



```
misc_feature      /note="assembly_name:Contig14"
                  1196. .2203
misc_feature      /note="assembly_name:Contig15"
                  2304. .9462
misc_feature      /note="assembly_name:Contig55"
                  9563. .25807
misc_feature      /note="assembly_name:Contig56"
                  clone_end:sp6"
                  25908. .39871
misc_feature      /note="assembly_name:Contig57"
                  39972. .68325
misc_feature      /note="assembly_name:Contig58"
                  68426. .138688
misc_feature      /note="assembly_name:Contig59"
                  30000 c 29735 g 37221 t
BASE COUNT      600 others
ORIGIN
```

Query Match	34.3%	Score 1952	DB 2	Length 138688
Best Local Similarity	100.0%	Prod. No. 0		
Matches 2002	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY 3677	ggcaagcaggaataatgcacagctggtctcccaagcattctggagcagaactactccc			
Db 136248	ggcaaacaggaataatgcacagctggtctcccaagcattctggagcagaactactccc			
QY 3737	ctgggaagagctctcagagtgctccaaagagagaaccccaaggtactccagctaccaggtctca			
Db 136188	ctgggaagagctctcagagtgctccaaagagagaaccccaaggtactccagctaccaggtctca			
QY 3797	ggaacggtctactctggagagcatgctctcaagccagaggtcatgctggaactcagagagc			
Db 136128	ggaacggtctactctggagagcatgctctcaagccagaggtcatgctggaactcagagagc			
QY 3857	tccttcgcacggagaaacagagcggaagggcgagcaatgaaagaaagcctctccgaag			
Db 136068	tccttcgcacggagaaacagagcggaagggcgagcaatgaaagaaagcctctccgaag			
QY 3917	ggcccgcaactctgactctgataagaagtccaaagaccacaggttaagccctcccaag			
Db 136008	ggcccgcaactctgactctgataagaagtccaaagaccacaggttaagccctcccaag			
QY 3977	ggcccttcggagaagatgctgcacctccacctccactcaggttggaagctgaaagacttc			
Db 135948	ggcccttcggagaagatgctgcacctccacctccactcaggttggaagctgaaagacttc			
QY 4037	agacccctggagaaagggaagccctctatctccagcagcaataaagagatgctcaatg			
Db 135888	agacccctggagaaagggaagccctctatctccagcagcaataaagagatgctcaatg			
QY 4097	tcgcgcaataaagacatttctctctgaagactgtaatttggaggtttttttaaaccct			
Db 135828	tcgcgcaataaagacatttctctctgaagactgtaatttggaggtttttttaaaccct			
QY 4157	cgatggtactaaggagatcttctgltgttgtatgaagtgacctttaaagcggctgaagcaa			
Db 135768	cgatggtactaaggagatcttctgltgttgtatgaagtgacctttaaagcggctgaagcaa			
QY 4217	gaaatggaaagccctaaatgctctgccaataatgctcaagtgctgttttcaatggaagat			
Db 135708	gaaatggaaagccctaaatgctctgccaataatgctcaagtgctgttttcaatggaagat			
QY 4277	ttcccaaccctgtgaacataatctgtttgaggtgtgaatagctctcgagcctccacagt			
Db 135648	ttcccaaccctgtgaacataatctgtttgaggtgtgaatagctctcgagcctccacagt			
QY 4337	acacgagatctcgccctactcaatgaatttccggagcttggtgtlaacatccctgaacc			
Db 135588	acacgagatctcgccctactcaatgaatttccggagcttggtgtlaacatccctgaacc			
QY 4397	acttcagatgaaacttactcaagtatgaggaatgaactgctgagccctggtttctccctca			

Db 135528 ACTTGCAATGACAAATTGACCTTAAGTGGAGATGACGTGGCCCTGTGTTCTCTCTCTA 135469  
 Oy 4457 agctcttagctatgagatgacatcttagctctggtggagaaagtggcgacatc 4516  
 Db 135468 AGTTCTCTTAGCTATGGATGATGACATCTTAGTCTCTGGGGAGGAAAGTGGGACATAC 4516  
 Oy 4517 accaaaatttggggcttcttgtaactcaagcaacaaatctgtcgaacttgcataca 135409  
 Db 135408 ACCAAAATTTGGGCTTCTGTGACTTTCACAGCACACCACTTGTGTACTTTGTCATCA 4576  
 Oy 4577 ctgtgttcttctcttctcttctgaactcttctgtcagggaagtcgtcactatcac 135349  
 Db 135348 CTGTGTTTCTCTTCTCTTCTTCTCAAGCTCTTGTGTGACGGGAAAGTGGGTCACTTAATAC 4636  
 Oy 4637 agaagctaaagcctatgccaacatcttctgtccacatgggggtctcttctgaactcct 135289  
 Db 135288 AGAAGCTAACCCATAGTCCACAACTCTTGTGTCACATGGGGGTCCTTTGTAACTGGCT 135229  
 Oy 4697 tatgcatacaatcattaccaataagtgatgattccgtgtctgtttatacctaagcttct 4756  
 Db 135228 TATGACTCAACATTTACCAATAAGATGATGATCTGTCGTCTGCTTATACATACGCTGT 135169  
 Oy 4757 cgtctcttctcctgaacgtgtgtgtgtgtgaaccaacagctctgtgtgtggaaactgtggaga 4816  
 Db 135168 CGTCTCTGTCTCTGACGTGGGTGTGATCCACACAGCTCTGTGTGGGAACGTGGGAGA 135109  
 Oy 4817 caaggtgtgcctgcctgcgggggaagctgggacctgtgcacttggcctgtctcatatgag 4876  
 Db 135108 CAGGAGTGGCTCCCTCCCGGGGGAAGCTGGGCTCCGCTATTTGGCTGTCTATCATGAGG 135049  
 Oy 4877 ggagaagctaaagaagaatctctcttaggaagaagctcatgtgccagatcacatcctaact 4936  
 Db 135048 GGAGAGCTTAAGAAAGAAATTCCTAGGAAGAGCTCAATGGCCAGTACATCTAGTAAT 134989  
 Oy 4937 attctaagttctgtctcagagcctgtcagaagaagcacagaatggagcagaagata 4996  
 Db 134988 ATTCTAATTAATGTTTGTGTTCTGACAGCTGTCTGAGAAAGGACACAGATGGACAAAGATA 134929  
 Oy 4997 aaccgaagcattatcttgaatctctctcctcaggttttctaagctcaagagcaatctgaatc 5056  
 Db 134928 AACCGACAGTCATTTGATCTGCTCTCTCAAGCTTTTCAAGTCAAGAGCAATTTGATGCT 134869  
 Oy 5057 tgtctaataagatccaacactgcatctcgtactggcgatgcacagctcctcctaagttct 5116  
 Db 134868 TGTCTAATCATCCACACTGCATGCTGATCTGGCGATGCCGCTCTTAAGTAGTCT 134809  
 Oy 5117 gccatgaacataaaagaagaagaagccgttacacatcaacagaagaacttctcg 5176  
 Db 134808 GCCATGAACATAAAAGACAAAGAAAGCCGTTACATCACACAGAAACAACTTTTCGG 134749  
 Oy 5177 gtcccaacacgtgtgtgccaagaaagctcactctcgtcgaagctttagagttgtgtcgg 5236  
 Db 134748 GTCGCCACAGCGGTGTGTGCGAAGCTCACCTCTCGGTCAGTATTAAGATGTGTGTGG 134689  
 Oy 5237 gctcggggaatcgtgtgtctccatctctctcactatgttctgaacatccgtatcttaa 5296  
 Db 134688 GTCGTGGGATCTGGGTGCGCTCCACTCTTCCTTCAATGTCTGAAACATCTGTAATTTAA 134629  
 Oy 5297 acctgtctgggtgtcctaagtggtcgtgtgaatcccgatgtggaanaagctgggggtgaaa 5356  
 Db 134628 ACCATGGCTGGGTGTAAAGTCTGTGAATCCGATGTGAAAAAGCTGGAGGGTAA 134569  
 Oy 5357 gtctgaatcacatctatcttcttaaaaaagaanaaagaacatctatgatgatgta 5416  
 Db 134568 GCTCACATACATGTAATTTACTTTAAAAACAGAAAAAAGACATATAGATATAGTCTA 134509  
 Oy 5417 tttttttttttggacatctgtaattcttctgtgtgtgttctttagaaatgagtgtc 5476  
 Db 134508 TTTTTTTTTTTGGCAATTTGTAATTTTGTGTGACCTGTTTTAGAAAAAGATGTCTC 134449  
 Oy 5477 acaagctaccctgtctctctcgcactctctgtcgaatgtctgtcttcttaataagct 5536  
 Db 134448 ACAAGCTACCCTGTCTCTCTCGCATTTCTGTGTGACAGGTCTGTTTCTTAATCAAGT 134389

QY 5537 cggcggttctaagtgtgttaccaggtgacgcagtgacgttgatgacagtgtgctt 5596  
|||||  
Db 134388 CGCGGTCGTCTAAGTGTGTGTACAGTGTACGCGAGTACCTGTGATGACAGTGTCTCT 134329  
|||||  
QY 5597 ttctcacagcctccctcagctgtgtgagaacagcttctctgtacatgcaactcctaa 5656  
|||||  
Db 134328 TTCTCAGAGCCCTCCCTGTGAGTGTGAAACAGCTTCTCTCTACATGACAACTCTTAA 134269  
|||||  
QY 5657 taaagcatttcttctctgtt 5679  
|||||  
Db 134268 TAAAGGCATATTCTCTCTGT 134246  
|||||  
RESULT 10  
AK024668 2116 bp mRNA linear PRI 29-SRP-2000  
LOCUS Homo sapiens cDNA: FLJ21015 f1s, clone CAE05730, highly similar to  
DEFINITION AK024668  
ACCESSION AK024668  
VERSION AK024668.1 GI:10437004  
KEYWORDS oligo capping; f1s (full insert sequence).  
SOURCE Homo sapiens primary endothelial cells of human coronary artery  
cDNA to mRNA, clone\_id:CAE clone:CAE05730.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Kawabata,A., Hiki,J.T., Kobatake,N., Inagaki,H., Ikema,Y.,  
Okamoto,S., Okitani,R., Ota,T., Suzuki,K., Obayashi,M., Nishi,T.,  
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
TITLE NEO human cDNA sequencing project  
JOURNAL Unpublished (2000)  
REFERENCE 2 (bases 1 to 2116)  
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.  
TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio  
Sugano, Institute of Medical Science, University of Tokyo,  
Laboratory of Genome Structure Analysis, Human Genome Center;  
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
Fax:81-3-5449-5416  
COMMENT NEO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction: 5'-6' 3'-end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).  
FEATURES  
source Location/Qualifiers  
1..2116  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="primary endothelial cells of human coronary  
artery"  
/clone="CAE05730"  
/clone\_id="CAE"  
/note="cloning vector pME18SFL3"  
1..2116  
/note="highly similar to AF252293 Homo sapiens PAR3 (PAR3)  
mRNA"  
BASE COUNT 651 a 476 c 571 g 418 t  
ORIGIN  
Query Match 33.4%; Score 1899; DB 9; Length 2116;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1899; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2227 gtaataacacagcttccctacagtgatgataatgacacagacgacgctgacattataga 2286  
|||||  
Db 200 GGTAAATACAGCTGTCCCTCTACAGTGAATATGCCCAAGATGACACTGTCTATTATGAA 259  
|||||

QY 2287 gatgacaggttgcagtgcttccctccacatctctctgacaggtcctcttccagctccat 2346  
|||||  
Db 260 GATGACAGAGTGTGGCAGAGCTCTCTCCACATCTCTGTACACAGATCTCTTCCAGTCCCAT 319  
|||||  
QY 2347 gatgatgtgtgggttctgtacagcagatgctgtgacttgggccaaggtctgacatgat 2406  
|||||  
Db 320 GATGATGTGGGGTGTGTGAGAGGAGATGCTGTACTGTGGGCCAAAGGCTGCATACAGTAT 379  
|||||  
QY 2407 tcaagcagctgctctttagtccagatgttgatccagttcttcttcttcaagagaagga 2466  
|||||  
Db 380 TCAGCCGACTGCTCTTGAATCCAGATGTGATCCAGTTCTCTTCAACAGAGAAGGA 439  
|||||  
QY 2467 ttggaagctcagagtatgtacaagaaacgcacaaagcaatttcaatgtgacatgcaatg 2526  
|||||  
Db 440 TTGTGACGTACAGATATGTACAGAAAAACGACAAAGCAATTTTCAGATGCCAGTCAATTG 439  
|||||  
QY 2527 gattcgttaaaacacagaaatacaaaaaagcatalgatttaagtalagctgacagactaa 2386  
|||||  
Db 500 GATTTCGTTAAAAACAGAAATCAAAAAGCATGGATTAGTATAGCTGACGAGACTTAA 559  
|||||  
QY 2587 ctaataacagtgatgataccagaaagcaggttctcccaagcagatgtgggtccttccctg 2646  
|||||  
Db 560 CTCAAATACAGTGGATGACCCAGAAAGCAGGTTCTCCACAGAGATGTGGGTCTTCCCTG 619  
|||||  
QY 2647 ggtctgaagaagtcacagctcgttggagagctgtcagacccagcttgcagagtgacttg 2706  
|||||  
Db 620 GCTCTCAAGAGTCACAGCTGTTGGAGAGTCTCAGACGCCCACTTGGCGAGGTGACTTTG 679  
|||||  
QY 2707 aatgggataatcttccttccatcgctcagcgccgagtaatacagagcagggatgacat 2766  
|||||  
Db 680 AATGGGGATATTTCCTTCCATCTCCACAGCCGCCGCGTAAATCAGAGGCAGGGGATGCAT 739  
|||||  
QY 2767 gaaagcttcagagctgcctcagcacaactcttatgataaacccgcggtgagatgatatg 2826  
|||||  
Db 740 GATAGCTTCGAGCTGCTCATCGACAAATCTTATGATTAACCCCGGTGATGATGATGAT 729  
|||||  
QY 2827 gaaagcattggaagccttggaaagaacacagaaagaagtcacagatcaggaagaagatc 2886  
|||||  
Db 800 GAAGCGATGAGAGCTGTGGAAAGACACAGAAAGAAAGTTCAAGATCAGGAGAGAGTCT 859  
|||||  
QY 2887 gttctcacagcagctgtatcagcttccctcctctctgtgagaacaaatgaatgaaaccaa 2946  
|||||  
Db 860 GATTCACACAGCCAGTATATCCCTCCACTCTCTGTGAGAGCAAAATGGAATGGAACCAA 919  
|||||  
QY 2947 gaaagaagtgataagactgataagaaagaagtaaaactgtgaaagaagaagaagaagat 3006  
|||||  
Db 920 GACAAAGGTGATTAAGACTGATTAAGAAAAGATTAAGTAAGTAAGTAAGTAAGTAAGTA 979  
|||||  
QY 3007 aagaataaggaagaagataaaatgaaagccaagaaggaaatgctgaaaggtcttggagac 3066  
|||||  
Db 980 AGAGATTAAGAGAAAGATTAATTAATGAAGCCAAAGAGGAATGCTGAAGGGCTTGGAGAC 1039  
|||||  
QY 3067 atgttcaaggttggcaaacatcgaaagaatgacaagaattgaaagaacgggtataataaaa 3126  
|||||  
Db 1040 ATGTTTACAGTTTGGCAAACTCGAAAGATGACAGTTTACAGAAACGGGTAAATAATAAA 1099  
|||||  
QY 3127 atacaggaatccttcatcagaagaaggaagatacgaatgaaagcagagcagaggaagga 3186  
|||||  
Db 1100 ATTCAGGAATCTTTCATTCATCAGAAAGAGAGAGATACGAATGAAGCAGAGCGAGAGAG 1159  
|||||  
QY 3187 attcaagcacaacatcgagaatlttgaagaaagcagacgtcgaagcgtgtgactatgtgaa 3246  
|||||  
Db 1160 ATTCAGCCCAAACTCGAGAAATTAAGGAAGCAGACAGCTCGAAGCGGTGACTATGCTGAA 1219  
|||||  
QY 3247 attcaagatttctatcggagacatttggctgtgatagagtaatgataatgtgagggagttct 3306  
|||||  
Db 1220 ATTCAAGATTTTTCATGCGACATTTGGCTGTGATGATGAGTATATGAGGGAGATTTCT 1279  
|||||  
QY 3307 tcttaagaaagttcacttggtctcgaagcgtcagaagccccaagcccaagaaagggcatg 3366  
|||||  
Db 1280 TCTTATGAAGTTCATGTGCTCTCAACGCTAGCTAGAGCCCAAGAGAGGGCATATG 1339  
|||||



757-781-20.01ig.rge

Page 22

COMMENT

Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA  
Note: Clone was sequenced in 1996 using Multiplex DNA Sequencing technology. Data may contain low quality sequence and BAC/contig vector sequences.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs is believed to be runs of N. The order of the contigs provided by the submitter are based on estimates that the sequence will be replaced by the finished sequence and BAC/contig  
\* the accession number as soon as it is available and I 22279: contig of 22279 bp in length.  
Location/Qualifiers  
1. 22279  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/map="10p11.2"  
/clone="CRI-JC2075"

BASE COUNT 5671 a 5238 c 5397 g 5959 t 14 others

ORIGIN

Query Match  
Best Local Similarity 31.0%  
Matches 2002; Conservative 0; Mismatches 0; Indels 2; Gaps 2.

3677 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
9469 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
3737 ctggaggaagcttcagagtcgcaagaagaccacaggtactcaagctaccgaagctcca  
9529 ctggaggaagcttcagagtcgcaagaagaccacaggtactcaagctaccgaagctcca  
3797 ggaagcgaagcttcagagtcgcaagaagaccacaggtactcaagctaccgaagctcca  
9589 ggaagcgaagcttcagagtcgcaagaagaccacaggtactcaagctaccgaagctcca  
3857 tccctgcagaagagagagcggaagagcagagtcacaggtcagtcgaggaactcaggagc  
9649 tccctgcagaagagagagcggaagagcagagtcacaggtcagtcgaggaactcaggagc  
3917 tccctgcagaagagagagcggaagagcagagtcacaggtcagtcgaggaactcaggagc  
9709 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
3977 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
9769 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
4037 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
9829 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
9888 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
4157 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
9948 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
4217 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
10008 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
4277 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc  
10068 ggcgaagcagaagaatgcagctcggtcccgagctctggagcgaactactccc



QY 4337 accgaatctcggccctactcatagatgtlccggcgttggtgttaacatccctgacc 4396  
|||||  
DB 10128 ACCAGAACTCGGCCCTACTCATGAGTGTGCGCGCTTGGTTGAACATCCCTGACC 10187  
QY 4397 acttgcagagacaaatctcccggaagtgagatgacgctggccctgttctccctcta 4456  
|||||  
DB 10188 ACTTGAGTGAACAAATTCACCTGAAGTGAAGTGAACGTGCGCCCTGTTCTCCCTCTA 10247  
QY 4457 agtctcttagctatgagatgacatcttagctctgtgtgagaaaaagtggcgacatac 4516  
|||||  
DB 10248 ACTTCTCTTAGCTATGSGATGACATCTTAGTCTCTGTGTGAGAAAAAGTGGCGACATAC 10307  
QY 4517 accaaaaattggggtcttctggtlaactcaagacaaagccattgtctgtaattgtcatca 4576  
|||||  
DB 10308 ACCAAAAATTTGGGCTTTCTGCTACTTTCACACAGCACAGCCATTGTCGTACTTGTTCATCA 10367  
QY 4577 ctgtgtttct 4636  
|||||  
DB 10368 CTGTGTGTTTCT 10427  
QY 4637 agaaagcctaagccatagtcacacatgttctgtcacaatgggggtcccttctgaactgctc 4696  
|||||  
DB 10428 ACAAGCTAAGCATAGTCCAAATGTTTGGTACACATGGGGTCTTTGTAACTGCTCT 10487  
QY 4697 tatgactcaacattaccaaataagtgatgacccctgtgtctgtctgttatacatacgtctgt 4756  
|||||  
DB 10488 TATGACTCAACATTAACCAATTAAGTATGATCTGTGCTGCGCTTATACATACGCTTGT 10547  
QY 4757 cggctcgtgtctctgacagtggtgtgagtcacacagcctctgtgtggaagagctggagaa 4816  
|||||  
DB 10548 CGGTCTCTGTCTGACACCTGGCTTGTGACACACACCTCTGTGTGGAACCTGGGACA 10607  
QY 4817 caggagtgtctctctcggggggaagctgggctggtccattgtgacctgtctatcatagag 4876  
|||||  
DB 10608 CAGGAGTGGCTCTCGCGGGGGAAGCTGGGCTGCGCATTTGGCCCTGTCTATCATGAGAG 10667  
QY 4877 ggaagagtaagaagaatctctctctaggaagagctcatagggccagttacatcctaagta 4936  
|||||  
DB 10668 GGAGAGCTAAGAAATAATTCCTTAGGAAGAGCTCATGCGCCAGTACATCCTAGAAT 10727  
QY 4937 atttaataatgttctgtct 4996  
|||||  
DB 10728 ATTTTAATTAATGTTTGTCTGACAGCTTGTGAGGAAGGCGACAGATGGGACAGAGATA 10787  
QY 4997 aaccagacagtcatttgaactgtctctctctctctctctctctctctctctctctctctct 5056  
|||||  
DB 10788 AACGAGACAGTATTTGATCTGCTCTAGCGTTTTCATAGTCAGAGGCAATTTGATGCT 10847  
QY 5057 tgtctaattgcatccacacacatgcatgtctgac-tggcgaatgccaagctctctaagtatgtc 5115  
|||||  
DB 10848 TGTCTAATGTCATCCACACACTCTCATGTCTGATGGCGATGCCAGCTCTCTAAGTATGTC 10907  
QY 5116 tggccttgaaacataaagaagaagaagaagccgttatacacatcacagaagaacatttctg 5175  
|||||  
DB 10908 TGCATGAAACATTAAGAAAGCAAAAGAAAGCCGTTTACATCACACAGAGAACATTTTGG 10967  
QY 5176 ggtcccaacagcggtgtgtgcaagaaagctcactcctcgctcagttatagagtgtgtgtg 5235  
|||||  
DB 10968 GGTCCACACGGGTGTGTGGCAGGAAGCTCATCTCGCGTCATATTTAGAGTGTGTGTG 11027  
QY 5236 ggtctcggggatctcgtgtgtcccatcttctctcaattgttcttgaaacatcctgtattgta 5295  
|||||  
DB 11028 GGTCTCGGGGATCTCGGTGGCTCCCATCTTTCCTCATTTGTCTGAAACATCTGTATTTGTA 11087  
QY 5296 aaccatgtctgggtgtgtctaaagtgtgtgtaatcccgatgttgaaaaagctggaggtgaa 5355  
|||||  
DB 11088 AACCATGCTGGGTGTCTAAATGCTGTGAATCCGATGTGAAAAAAGCTGAGGTGAA 11147  
QY 5356 agctcagcatcacatgtattctcttaaaaaacagaaaaaacatgtgtgtgatatgtct 5415  
|||||  
DB 11148 AGCTAGCATACCATGTATTTACTTTAAACAGAAAAAAGAACATGTATGATATGTCT 11207

QY 5416 attttttttttatggcacattgattttgtgttgactgtttttagaaatgatgtgtc 5475  
|||||  
DB 11208 ATTTTATTTTATTTGGCACATTTGATTTTGTGTGACTGTTTGTGAATGATGTGTC 11267  
QY 5476 cacacagtaacccgtgtct 5535  
|||||  
DB 11268 CACACAGTACCCGTGCTCTCTGCACTTCTGTGTCATGATGTTCTGTTTAAATCAAGT 11327  
QY 5536 gcgcggtgtcttaagtggtttacacagtgtaacgcagtgacttgatagatgagtgctc 5595  
|||||  
DB 11328 GCGGCGGTGCTTAAGTGTGTATTACAGTGTACCGGAGTACCTTGATGACAGTGGCTC 11387  
QY 5596 ttcttcaagctcctccctgagctgtgtgagaaagccttctctgtlaacatgacaactccta 5655  
|||||  
DB 11388 TTTCTCACACCTCCCTCGAGCTGTGAGAAACAGCTTTCCTGTACATATGCAATCTCTA 11447  
QY 5656 ataaagacatattctctctgt 5679  
|||||  
DB 11448 ATTAAGGCAATTTCTCTGTT 11471

## RESULT 12

LOCUS BD004905 2718 bp DNA linear PAT 31-JAN-2002

DEFINITION Novel genes encoding protein kinase or protein phosphatase.

ACCESSION BD004905

VERSION BD004905.1 GI:18632866

KEYWORDS JP 03074935-T/2.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K., Otsuki,T., Funahashi,S., Seno,C. and Nezu,J.

TITLE Novel genes encoding protein kinase or protein phosphatase

JOURNAL Patent: JP 03074935-T 2 30-JAN-2001; HELIX RESEARCH INSTITUTE,TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA, KOJI HAYASHI,KAORU SAITO,JUNICHI YAMAMOTO,SHIZUKO ISHII, OMOYASU SUGIYAMA, AI WAKAMATSU,KEIICHI NAGAI,TETSUJI OTSUKI,SHINICHI FUNAHASHI, HIKI SENO, JUNICHI NEZU

COMMENT OS Homo sapiens (human)

PN JP 03074935-T/2

PD 30-JAN-2001

PF 28-JUL-2000 JP 200005061

PR 29-JUL-1999 JP 99P 248036,11-JAN-2000 JP 00P 118776 PR

02-MAY-2000 JP 00P 183767,09-JUN-2000 JP 00P 241899 PR

18-OCT-1999 US 60/159590,17-FEB-2000 US 60/183322 PT TOSHIO

OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI, PI KAORU SAITO, PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU, PI KEIICHI NAGAI,TETSUJI OTSUKI,SHINICHI FUNAHASHI,CHIAKI SENO, PI JUNICHI NEZU

PC C12N15/12,C12N9/12,C12N9/16,C12N1/15,C12N1/19,C12N1/21, PC

C12N5/10,C07K16/40,

PC C12Q1/68,C12Q1/42,C12Q1/48

CC

FH Key Location/Qualifiers

FT CDS (33) (2627).

FEATURES

source 1..2718

BASE COUNT 840 a 586 c 688 g 604 t

ORIGIN

Query Match 29.8%; Score 1694; DB 6; Length 2718;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1844; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 374 ttgaagtcacaccttaagctctctcgagcaaatatgctcttcaatgtgttgagcagtagtg 433

OY	1514	ttaaggcaggaagaacactatagaagtcaattgtagttagtttagggccaatccaaag	1573
Db	1141	TTTAAGGCAAGNACAGACTTATAGAGCTAAATGTGGATGATTATAGTGGAATAATGCCAAC	1200
OY	1574	aggaagtgttttcgcgttgttgagaagcaccaagaatggaaaggaactctgtagccttcgtctc	1633
Db	1201	AGGAAGCTGTTTCGCCTGTTGAGAAGCACCMATGAAAGAACTGTAGCCCTTCGTCT	1260
OY	1634	ttcgccagaagaagacgccctccaccsaagggaactgaatgcagagcccaaacgcagatgcaga	1693
Db	1261	TTTCGCCAAGAAAGAGCGCCTTCCACCACAGGAACTGAATTCGACAGCCMACCCAGATGCACA	1320
OY	1694	ttccaaaagaagaacgaagaacagaagaataaggatalgttcttcaccctgatgtgaccaggg	1753
Db	1321	TTTCCAAAAGAAACGAAACGACAGATGAGATTTGTTCTTACACCTGATGGCACCAAGGG	1380
OY	1754	aattcttcgactttgaagtcaccaactaatgatctcagatctgcgaagcgctgtgtcagtg	1813
Db	1381	AATTTCTBACATTTWGAAGTCCCATCTTAGATGATTCAGATCTGCAGCGCTTGGGTCTCATG	1440
OY	1814	tcaaaggttaacccgttcaaaagaagaacaacgacagattggnaatctcttgtaagtcatta	1873
Db	1441	TCAAAGGTAAACCGGTCCMAAAGAAAGAACCCACACATTTTGGGAATCTTTGTCAAGTCATTA	1500
OY	1874	tttatggaggagacagcatctaaagaatggaaagcttcgggtgaatgatgcacatgatagcag	1933
Db	1501	TTAAATGGAGGAGCAGCATCTAAAGATGAGMAGCTTTCGGGTGAATGATCAACTGATAGCAG	1560
OY	1934	ttaatggagaatcccgttggtggcagaacaacsaagatgcacatgaaaccttaagaagct	1993
Db	1561	TAAATGGAGATCCCTGTTGGGCAAGACAAACCAATCCATGTAAGAAAGGT	1620
OY	1994	ctatgcttactaaagcaataaacaggaatatccagcttattgtgtgcaagaagaataa	2053
Db	1621	CATGTGCTACTGAAAGCAATPAACGAGGAATATCCACCTTAATGTTGCCAKMGAGAAATTA	1680
OY	2054	ggaagtgaagagactggaagtcacccctgggaagcccccttgtaactgtgactgtgccattgaa	2113
Db	1681	GCAAGTGCATGAGCTGAAAGTCACTCGTGGAGCCCCCTGTGACCTGAGCTGCACATTTGAAA	1740
OY	2114	cagcgttgaatgatagagaagaagaattccattccctcctacagtgaggatgtgaggggc	2173
Db	1741	CACCGTWTGATGATAGAAACAGAAATTTCCATTTCCCTCTACAGTGGGATTTGAGGGGC	1800
OY	2174	ttgatgaatgcgccagcagaatgctgcacctcaagtagaataatgggt	2220
Db	1801	TTGATGATGATGCCACGAGAAATGCTGCCTCTCACTAGATTAATGGGT	
RESULT_13			
AKO27735			
LOCUS			
DEFINITION	Homo sapiens CDNA FLJ114829 fls, clone OVAC1000945, partial		
ACCESSION	AKO27735		
VERSION	AKO27735.1		
KEYWORDS	oligo capping; fls (full insert sequence).		
SOURCE	Homo sapiens ovary; tumor tissue cDNA to mRNA, clone_11b:OVARC1		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Nishikawa,T., Ota,T., Hayashi,K., Sugano,S., Sugiyama,T., Otsuki,T., Suzuki,Y., Megatsuna,M., Hosokiri,T., Kaku,Y., Kodaira,H., Sudo,H., Ono,Y., Takiguchi,S., Wakabe,S., Kimura,K., Murakawa,K., Nakamura,Y., Nagaba,Saito,K., Yamamoto,K.,		
TITLE			

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2718)  
AUTHORS Isogai, T. and Otsuki, T.  
TITLE Direct Submission  
JOURNAL Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 299-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3951, Fax: 81-438-52-3952)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing,  
Research Association for Biotechnology; cDNA library construction,  
5'-6' 3'-end one pass sequencing and clone selection; Helix  
Research Institute (supported by Japan Key Technology Center etc.)  
and Department of Virology, Institute of Medical Science,  
University of Tokyo.

FEATURES  
source Location/Qualifiers  
1..2718  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="OVARC1000945"  
/tissue\_type="ovary, tumor tissue"  
/clone\_id="OVARC1"  
/note="Cloning vector: pME18SFL3"  
33..2630  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB55330.1"  
/db\_xref="GI:14042634"  
/translation="MPLHVRSSDPALIGLSTYSVDSNFSSEPSRKNPTNMTTAF  
LKONTASPKACDRKDENYRSLPRDTSNMGNOFORNDRSVNAGLATVQAPRLNH  
ODEDGEDNDRVRVPHADNGLEHINPEISLDVAVKLVENDDGPGILHVPFARC  
GRTIGLVKRLKESGKAHENVLPENDCIYVINGDLNRRNPEQAHFRAMRPTII  
WFHVVPANKEQYEQLSQSEKNNTYSSRFSPDSYIDNRSVNAGLATVQAPRLNH  
PEQIDSHRLPHSHPGKPSAPASAPQNVSTSVSGYTKRIGKLNLOLKTE  
GLGFSITSDVTLIGSAPLYKNIILPRGAALODRKLKAGDLIVNGVDLYGQSEV  
VSLRSTKMEGTVALVLRQDAHPRELNPQMPKTKADEDIVLPQSTR  
FLFEVPLSDSGSLGVSVKSRKSNHNLGIFFVSIINGASAKGRIRVNDLI  
AVNGESLIGKTNQDAETLRKSMTEGKRKMIDLIYARRISKNEIKSPSPPEPL  
PIETALDDREKRISHLSYSGIEGLDESFRNALSRIMGKVLSPVAMPDDVYIIE  
DRLPLVPHLSDSSSSSHDDVGFVADACTAKAALISDADCSLSPDVPVAFOR  
EGFGRQDETLKLVNDQKAGSPRDVGPSTGLKSSLSLQTAVEVTLNGDIPFI  
RPRRIIRGRGNCNFRALIDKSYDKPAVDDDGEMETLEEDTESSRSRGSEVSTAS  
DPSHSLEORONGNDEKDKTRKKDKTKRKKRDKRKKAKKMLKGLDMFS  
LAKLKEPRK"

BASE COUNT 840 a 586 c 688 g 604 t  
ORIGIN

Query Match 29.8%; Score 1694; DB 9; Length 2718;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1844; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 374 ttgaggtacacacctcaagtccttcagcaaatatgctcttcattgctcgagcagtaagtg 433  
|||  
Db 1 TTGAGGTACACACCTTCAGTCCCTTCGAGCAAAATATGCCCTTCATATTTGAGCAGCAATG 60  
|||

QY 434 acccagcttaattggtccttcacactctgtcagtgatgataatttccttccttgaagagc 493  
|||  
Db 61 ACCAGGCTCTATTGGCTCTCCACTTGTCTAGATAGTAATTTTCTCTGAAAGAC 120  
|||

QY 494 ctcaagaaaaatccacacagctgtgtcaacacacagctgctctcctcaagcaagacagtg 553  
|||  
Db 121 CTTCAAGAAAAATCCACACGCTGTGTCAACACAGCTGGCTTCTTCAAGCAACACTG 180  
|||

QY 554 ctggagagtcctlaaaacctgagcagagaaagatgaaaaatcacacacagcctcccgcgag 613  
|||  
Db 181 CTGGAGTCTCTAAACCTGCGCAGAGAAAGATGAAGAACTACAGAACCTCCCGCGG 240  
|||

QY 614 atactagtaactgtgtctaaccaatttcagagagagaatgctgcgtctcctgaatgcca 673  
|||  
Db 241 ATACTAGTAAGTGTGTCTAACCAATTTTCAGAGAGCAATGCTGCTGCTCTGATGCCA 300  
|||

QY 674 gtcaaccatgtgtggcaagtgctggaagaaacagaaatgagatgagtggaagag 733  
|||

Db 301 GTCACCCAATGGTGGGCAAGTGGCAGAGACAACAAACAGATGAGATGGCAGCAAG 360  
|||  
QY 734 aggataacagtcgtgtgtgaacctgtgtgacatgtctgacaggglttgagacatatacca 793  
|||  
Db 361 AGGATTAACAGTGTGTGTGAACCTGTTGACATGCTGACACGGGTTTGAGACATATACCA 420  
|||

QY 794 actttctctgatatgatatgttaagctcgtagaagtcctcccaagatgagggcctctgg 853  
|||  
Db 421 ACTTTCTCTGTGTGTATGTGAACCTGTGTGMACTGTCCCAACGATGAGGGCCTCTGG 480  
|||

QY 854 gaatcatatgtgcttccatcagtcgcagagcgaggaacccctggggttattagtaaac 913  
|||  
Db 481 GAATCATATGATGCTTTCAGTGTGAGGCGGCAACCCCTGGGTTATTAATAAAC 540  
|||

QY 914 gattgagaaagggtgtgtaaaagctgtaacatgaaatcttltcgtggaatgattgcatg 973  
|||  
Db 541 GATTGAGAAAGGTGTGAAGCTGAACATGAAATCTTTTGTGTGAATGATGATGATG 600  
|||

QY 974 tgaagataatgattggtgagacctcgaaatagaagatttgaacagcaacatagtcttc 1033  
|||  
Db 601 TGAGGATTAATGATGGCGACCTTCGAAATAGAAGATTTGACAAACATATGTTC 660  
|||

QY 1034 gccaaagcctgctgtaacccatcatttggttccatgtgttctcgtgagaaataaagaagc 1093  
|||  
Db 661 GCCAAGCCATGGCTACACCCATCATTTGGTCCATGTGGTCTCGACAAATPAAAGAGC 720  
|||

QY 1094 agtatgaacacatatacccaagtgagaagaaacaattactatataaagcgtlttaagccgt 1153  
|||  
Db 721 AGTATGAACACATATCCCAAGTGAAGAACAATTAATTAATTAACCCGTTTAAAGCCGT 780  
|||

QY 1154 acaagcagatatattgcaaacagagatgtgaacagtgcaaggtctcaacgggtgcaagag 1213  
|||  
Db 781 ACAGCCAGATATTGTGAACAAGAGGTGTGACAGTCCAGGGCTTCCACAGGTGCAAGAG 840  
|||

QY 1214 caaccagcattgaacccacccgctgagcagataagcctcaactaagaactaacctcataagc 1273  
|||  
Db 841 CACCCGATGAACCAACCCGCTGACAGATTAACCTCTCACTAACAACATACCTCATAGGC 900  
|||

QY 1274 caacacccctcgaggaaaaacacacacacacccgctcagcctccgagcaccctcagaatgta 1333  
|||  
Db 901 CACACCCCTCGGGAANAACCATCCGCTCCACCTCGGAGCTCGAATGATTTAGTA 960  
|||

QY 1334 cgaactgaagcagtggttataacacacaaaaataggaagagcttaataatcagctta 1393  
|||  
Db 961 CGACTGTAAGCAGTGGTTTAACACCAAAAAAATAGGCAAGAGGCTTAATATCCACTTA 1020  
|||

QY 1394 agaaagtgtaagaagggtttggagattcagcatcactccagaagatgtaaaatagtggtc 1453  
|||  
Db 1021 AGAAAGTACAGAAAGGTTTGGATTCAACATCTTCCAGATGTGAACATATGTTGGCT 1080  
|||

QY 1454 cagctccaatctatgtgaaaaaatctcctcccgaggggcgcgcaaltcagaatgagccgac 1513  
|||  
Db 1081 CAGCTCAATCTATGTGAATAAATCTCTCCCGGGGGGGCGCATTCAGAGATGGCGAC 1140  
|||

QY 1514 ttaagggcaggaacagacttataagagtgaaatgagtagattagtgggcaaatcccaag 1573  
|||  
Db 1141 TTAAGGCAAGAAACAACATTAATAGAGTAATGAGATTAATTAAGGCGCAATCCCAAG 1200  
|||

QY 1574 aggaagtgcttgcgtgtgagaagcacaagaatggaaggaactgtggaacctctgtct 1633  
|||  
Db 1201 AGGAAGTGTTCGCTGCTTGAGAGACACCAAGATGGAAGAACTGTGAGACCTTCTGTCT 1260  
|||

QY 1634 ttgcgcaggaagacgctctccacccaagggaactgaaatgacagaccaaagcagatgcaga 1693  
|||  
Db 1261 TTCGCAGGAAGACGCTTCCACCCAAAGGAAGTGAATGATGACAGCCAGATGCAGA 1320  
|||

QY 1694 ttccaaagaaacgaaagaaagatgagatattgttcttacacctgtatggaccaaagg 1753  
|||  
Db 1321 TTCCAAAAAGAAAGAAAGAGAGATGAGATATTGTTTCACTGATGACACAGG 1380  
|||

QY 1754 aattctgacattgaagcccaacttaagatctcaagatctcgaagccttggtcagtg 1813  
|||  
Db 1381 AATTCTGACATTTGAAGTCCCACTTATGATTCAGAGATCTCAGGCTTGGTGTACGTG 1440  
|||

RESULT	14	
LOCUS	AF332592	
DEFINITION	Homo sapiens atypical PKC isoform-specific interacting protein	
ACCESSION	AF332592	
VERSION	AF332592.1	
KEYWORDS	GI:14579308	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLE	Fang,C.M. and Xu,Y.H.	
JOURNAL	Down-regulated expression of atypical PKC-binding domain deleted	
MEDLINE	Cell Res. 11 (3), 223-229 (2001)	
PUBMED	21520080	
REFERENCE	11642408	
AUTHORS	2 (bases 1 to 2967)	
TITLE	Fang,C. and Xu,Y.	
JOURNAL	Exon/Intron Structure And Splicing Variants of A Novel Human	
REFERENCE	Polarized Gene, hasip	
AUTHORS	Unpublished	
TITLE	3 (bases 1 to 2967)	
JOURNAL	Fang,C. and Xu,Y.	
REFERENCE	Submitted (29-DEC-2000) Laboratory of Molecular and Cellular	
AUTHORS	Oncology, Institute of Biochemistry and Cell Biology, Shanghai	
TITLE	Institutes for Biological Sciences, Chinese Academy of Sciences,	
JOURNAL	320 Yue-Yang Road, Shanghai 200031, P. R. China	
REFERENCE	Location/Qualifiers	
AUTHORS	1..2967	
TITLE	//organism="Homo sapiens"	
JOURNAL	//db_xref="taxon:9606"	
REFERENCE	//chromosome="10"	
AUTHORS	//map="10p11.2"	
TITLE	1..2967	
JOURNAL	/note="ASIPSB, alternatively spliced, similar to GenBank	
REFERENCE	Accession Number AF196186"	
AUTHORS	/codon_start=1	
TITLE	/product="atypical PKC isoform-specific interacting	
JOURNAL	protein short variant b"	

Db 1 ATGAAAGTGCACCGGTGTCCTGCGAGAGCCCGGGTGTGTCGCTGCGCTGCGGGAAGGCCAC 60  
 Oy 61 atgaagttcttcagcctcctcaacgaacgagtgatcccgctacccgaagagccatccgcaag 60  
 Db 61 ATGAAACTTTCAGCCTCATCCACACAGCGGTGACCCGCTACCGGAAGGCCATCCCAAG 120  
 Oy 121 gattccaaactactgatatacagatgacatcgcttggaaacaatgagaatgagaatctactagac 120  
 Db 121 GATCCAAACTACTGATATACAGTGCATCGCTTGGAACATGGAAATGAGGAATCTAGAC 180  
 Oy 181 ctgtatgcatctctctgtgtatgtaacagacagataaagacagactgtatagacagttgat 180  
 Db 181 CTGTATGCATCTCTTGTGTATGTAGCAGACGATTAAGACAGACTGTATGACAGTGTGTAT 240  
 Oy 241 gacagatcacatccatccacggagtgatgagacaggtgacgttccagttccagatgaccagaagc 240  
 Db 241 GAGAGAGATCCACATCCACGAGGAGTATGGCACCAATGCGCATTTCCAGGGATACCCACAGC 300  
 Oy 301 ccagagataatttggtagtgagcttgagacaaaatgtctcagcctttcagccttaccac 300  
 Db 301 CCAGAGATATTTGGTAGTGTAGCTTGACCAACAAATGTCCAGCCTTTCAACCCCTAACAA 360  
 Oy 361 gaaacaagtaaatgtgagttcacaccttcagtccttcagagcaaatatgctcttccatgt 360  
 Db 361 GCAACAACTAAATGTAGAGTACACACTTCAAGTCTCTTCAGCAAAATATGCTCTTTATGTT 420  
 Oy 421 cgaagcagatgtgacccagctcctaattggcctctccactcttccttgatagatgaatatt 420  
 Db 421 CGAGCGAGATGTAGACCCACGCTTAATTGGCTCTCCACTCTCTCAAGTAAGAATAATTTT 480  
 Oy 481 tcccttgaagagccttcaagaaatcccaacagctggttcaaaacagctggcttccc 480  
 Db 481 TCCCTTGAAGAAGCTTCAAGAAAAAATCCACACAGCTGGTTCMAACAAAGCTGGCTTCC 540  
 Oy 541 aagaagaacactgctggagatccaaaccttgagcagaagaagaatgaaacacaga 540  
 Db 541 AAGCAGAAACACTGCTGGGAGTCTCTAAACCTCGACAGAGAAATATGAAATCAACGA 600  
 Oy 601 agcctcccgaggactagttaacttgcttaacaatttcagagagacaatgagctgctg 600  
 Db 601 AGCCTCCCGGAGTACTAGTAACTGAGTCTTAACCAATTTCAAGAGACAATGCTCCCTCG 660  
 Oy 661 tctctgagtgccagtaaccacatggttgagcaagtgtgaggaacaagaacagatgag 720



```

Db      1 ATGAAATGACCGTGTGCTTCGGAGGAGACC66GTGTGTCGCCGTGCGGGAAGCCAC 60
Qy      61 atgaagtttcagctcatccagcagcggtgaacccgctacccggagggccatcgccaa 120
Db      61 ATGAAAGTTTTCAGCCTCATCCAGGAGGCGGTGAGCCCGCTACCGGAAGGCCATCGCAAG 120
Qy      121 gacccaactatgatatacaggtgcatcgcttggaacatggaatggagaaactactagac 180
Db      121 GATCCAACTACTGGATACAGGTGACATCGCTGGAAACATGGAAATGAGAGAACTATACAC 180
Qy      181 ctgagtcacatcttctgtatgtagcagcagataaagaacagctgtgacagtgctgac 240
Db      181 CTTGATGACATCTTTGTGATGATGACGAGACGATTAAGACAGCTGTGTGAGCTGTGAT 240
Qy      241 gagcagatccacatccagcaggtgtagcaccagtgccaggtccacgggttaaccagagc 300
Db      241 GAGCAGATCCACATCCAGGAGGTGAGCAGCAGTGCACATGTCACGGGTGACCCAGAGC 300
Qy      301 ccagaagatttgtagagcttgccaccaaatgtctacgcttcagccttacccaa 360
Db      301 CCAGAGATTTTGTGTGATGAGCTTGACCAACAAATGCTCAGCCCTTCAGCCTTACCA 360
Qy      361 gcaacaagtgaattgagtcacacactcagcttcagcagcaataatgctcttcatgtt 420
Db      361 GCAACAAGTGAATTTGAGGTACACCTTCAGTCTTCGAGCAAAATGCTCTTATGTT 420
Qy      421 cgaacgaatgagccacgctcctaatgtgctcctcactctctgtagtgaatattt 480
Db      421 CCAGCAGATGATGACCCAGCTTAATTTGGCTTCACATCTTGTGTGATGATGATTAAT 480
Qy      481 tccctgagaagccttcaagaaatccacacgctgtgcaaacacgctggtctctc 540
Db      481 TCCCTGAGAAGCCTTCACAGAAATCCACACGCTGTGTAACAAAGCTGGCTTCTC 540
Qy      541 aagcagacactgctgaggtcctaaacctgagcaggaagaagaatgaaactacaga 600
Db      541 AAGCAGAACCTGCTGGAGTCTCTTAAACCTGCGACAGAAAGATGAAACCTACAA 600
Qy      601 agcctccgggagactactgaactggtcctaaccaattgaagagacaatgctcgtcg 660
Db      601 AGCCTCCGGGAGTACTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 660
Qy      661 tctctgagtgccagtcacccaatgtggaagtgtgagagagaaagaagaatag 720
Db      661 TCTCTGAGTGCAGTCAACCCATGCTGGCAAGTGTGAGAAACAGAACAGATAG 720
Qy      721 gatggagagaagaagaatagtcgttgaacctgtgagcagtcgacagggcttg 780
Db      721 GATGGAGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
Qy      781 gagcatatcccaacttctctgtagatgatagtgaaagctcgtagaagtcaccaagat 840
Db      781 GAGCATATACCAACTTTCTCTGATGATGATGATGATGATGATGATGATGATGAT 840
Qy      841 gaggagcctctggaatccatgtagtgccttcagtgctcaggcggcagaaacccctgg 900
Db      841 GGAGGGCCTCTGGAAATCATGTAGTGTCTTCACTGCTGAGCGGCAAGACCTGGGG 900
Qy      901 ttattagtaaaacgattggaagaagtgtgaaagctgaaacatgaaatcttttcgtgag 960
Db      901 TTATTAGTAAACATTTGAGAAAGGTGTAAGCTGAAACATGAAATCTTTTCGTGAG 960
Qy      961 aatgattcattgtcaggaataatgtaggcgacctcgaaatagaagaattgaaacaagca 1020
Db      961 AATGATTGCAATGTCAGAGATTATGATGCGACCTTCGAAATAGAAATTTGAACAAGCA 1020
Qy      1021 caacatagttctgcaagaacatgctgtaaccacatcatttggtccatgtgtcctgca 1080
Db      1021 CAACATATGTTTGGCAACCCATGCTACACCCATCATTTGTTCCATGTGTGTTCTTGA 1080
Qy      1081 gcaaatagaagagatgaacaactatcccaagtgagaagaacaattactatccaagc 1140
Db      1081 GCAAAATTAAGAGCAGATGATGAACTATCCCAAGTGAAGAACTATTAATTAAGC 1140

```

Search completed: July 24, 2002, 14:33:32  
 Job time: 19019 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:33:29 ; Search time 21.33 Seconds

(Without alignments)  
6108.633 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 1356

Sequence: 1 MKVTVCGRFRVVPVPCGDGH.....SQVARLNRLQTPKGRPPYS 1356

Scoring table: OLIGO Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR\_71:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	6.5	1337	2 T13948	atypical protein k
2	11	0.8	1464	2 T13716	bazooka gene prote
3	11	0.8	2450	2 S71625	protein-tyrosine-p
4	9	0.7	1693	2 S76086	beta transducin-11
5	8	0.6	88	2 T15027	hypothetical prote
6	8	0.6	107	2 T39293	hypothetical prote
7	8	0.6	128	2 C86179	hypothetical prote
8	8	0.6	153	1 D71357	probable cation-ac
9	8	0.6	219	2 E84117	hypothetical prote
10	8	0.6	283	2 C64379	hypothetical prote
11	8	0.6	317	2 T27179	hypothetical prote
12	8	0.6	318	2 T24509	hypothetical prote
13	8	0.6	348	2 I49262	hypothetical prote
14	8	0.6	395	2 AD0380	multidrug efflux p
15	8	0.6	450	2 G01158	tyrosine kinase ac
16	8	0.6	475	2 I38597	zinc finger protei
17	8	0.6	592	2 S43570	C0385.6 protein (c
18	8	0.6	618	2 AB0186	probable exported
19	8	0.6	623	2 H87127	conserved hypochet
20	8	0.6	720	2 A45436	synapse-associated
21	8	0.6	721	2 D87753	protein C43E11.6 l
22	8	0.6	723	2 I38412	receptor tyrosine
23	8	0.6	724	2 JH0800	postsynaptic densi
24	8	0.6	737	2 S44862	R05D3.2 protein -
25	8	0.6	767	2 T09599	postsynaptic densi
26	8	0.6	852	2 T10811	channel associated
27	8	0.6	852	2 A72343	hypothetical prote
28	8	0.6	870	2 G01974	channel associated
29	8	0.6	882	2 I38912	receptor tyrosine

30	8	0.6	890	1 A53743	protein-tyrosine k
31	8	0.6	904	2 I38757	homolog of Drosoph
32	8	0.6	911	2 I56552	synapse-associated
33	8	0.6	926	2 I38756	homolog of Drosoph
34	8	0.6	1081	1 A42399	isolucine--trNA 1
35	8	0.6	1149	2 T27567	hypothetical prote
36	8	0.6	1281	2 T00346	hypothetical prote
37	8	0.6	1333	2 E84601	probable retroelem
38	8	0.6	1551	2 AB2410	WD-repeat protein
39	7	0.5	58	2 S75890	hypothetical prote
40	7	0.5	79	2 D69309	conserved hypochet
41	7	0.5	84	2 A13640	hypothetical prote
42	7	0.5	96	2 D64695	type I restriction
43	7	0.5	99	2 T35967	hypothetical prote
44	7	0.5	104	2 T28823	hypothetical prote
45	7	0.5	104	2 H72553	hypothetical prote

## ALIGNMENTS

RESULT 1  
T13948  
atypical protein kinase C isotype-specific interacting protein ASIP - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C/Accession: T13948  
R/Sum1, Y.; Hirose, T.; Tamai, Y.; Hirai, S.; Nagashima, Y.; Fujimoto, T.; Tabuse, Y.  
J. Cell Biol. 143, 95-106, 1998  
A/Title: An atypical PKC directly associates and colocalizes at the epithelial tight  
A/Reference number: Z17827; MUID:98437350  
A/Accession: T13948  
A/Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: mRNA  
A/Residues: 1-1337 <I2D>  
A/Cross-references: EMBL:AB005549; NID:q3868777; PIDN:BA034216.1; PID:q3868778  
C/Genetics:  
A:Gene: asbp

Query Match 6.5%; Score 88; DB 2; Length 1337;  
Best Local Similarity 100.0%; Pred. No. 8.9e-82;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 566 PKETRAEDIEDIVLPDGRREFLPPEVPLNDGSGAGLVGYNRSKRNHADLGIFVKSII 625  
DB 566 PKETRAEDIEDIVLPDGRREFLPPEVPLNDGSGAGLVGYNRSKRNHADLGIFVKSII 625  
QY 626 NGGAASKDGRRLVNDQLIANGESLGLK 653  
DB 626 NGGAASKDGRRLVNDQLIANGESLGLK 653  
RESULT 2  
T13716  
bazooka gene protein - fruit fly (Drosophila melanogaster)  
C/Species: Drosophila melanogaster  
C/Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C/Accession: T13716  
R/Kuchinke, U.; Grave, F.; Knust, E.  
submitted to the EMBL Data Library, November 1998  
A/Description: Control of spindle orientation in Drosophila by the Par-3-related PDZ-  
A/Reference number: Z17708  
A/Accession: T13716  
A/Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: mRNA  
A/Residues: 1-1464 <KUC>  
A/Cross-references: EMBL:AJ130871; NID:e1363519; PID:e1363520; PIDN:CA10224.1  
C/Genetics:  
A:Gene: bazooka  
A/Cross-references: FlyBase:FBgn0000163  
A/Map position: X



Query Match 0.8%; Score 11; DB 2; Length 1464;  
 Best Local Similarity 100.0%; Pred. No. 0.062;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 885 KKSSSESLQT 895  
 |||||  
 Db 1081 KKSSSESLQT 1091

## RESULT 3

S71625  
 protein-tyrosine-phosphatase (EC 3.1.3.48) RIP - mouse  
 M:Alternate names: epidermal growth factor-binding protein; serine proteinase  
 C:Species: Mus musculus (house mouse)  
 C:Date: 27-Nov-1997 #sequence, revision 12-Dec-1997 #text\_change 20-Jun-2000  
 R:Accession: S71625; S67987  
 R:Chida, D.; Kume, T.; Mukoyama, Y.; Tabata, S.; Nomura, N.; Thomas, M.L.; Watanabe, T.  
 FEBS Lett. 358, 233-239, 1995  
 A:Title: Characterization of a protein tyrosine phosphatase (RIP) expressed at a very early stage in mouse embryonic fibroblasts  
 A:Reference number: S71625; MUID:95145716  
 A:Accession: S71625  
 A:Molecule type: mRNA  
 A:Residues: 1-2450 <CHI>  
 A:Cross-references: EMBL:DB3966; NID:91232103; PIDN:BA12158.1; PID:91232104  
 R:Experimental source: strain DBA/2; cell line MEL 745A  
 R:Wolf, B.B.; Brown, M.D.  
 FEBS Lett. 376, 177-180, 1995  
 A:Title: Epidermal growth factor-binding protein activates soluble and receptor-bound src tyrosine kinase  
 A:Reference number: S67987; MUID:96105375  
 A:Accession: S67987  
 A:Molecule type: protein  
 A:Residues: 1098-1102 <WOL>  
 A:Experimental source: submaxillary glands  
 C:Genetics:  
 A:Gene: Ptpn13  
 A:Map position: 5  
 C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology;  
 E:566-860/Domain: protein 4.1 membrane-binding domain homology  
 E:1089-1165/Domain: GLGF domain homology <GLG1>  
 E:1361-1437/Domain: GLGF domain homology <GLG2>  
 E:1495-1574/Domain: GLGF domain homology <GLG3>  
 E:1769-1840/Domain: GLGF domain homology <GLG4>  
 F:1863-1937/Domain: GLGF domain homology <GLG5>  
 F:2203-2422/Domain: protein-tyrosine-phosphatase homology <PTP1>  
 F:2374/Active site: Cys (phosphotyrosine intermediate) #status predicted  
 F:2380/Binding site: substrate phosphate (Arg) #status predicted

Query Match 0.8%; Score 11; DB 2; Length 2450;  
 Best Local Similarity 100.0%; Pred. No. 0.099;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 502 DGRLLKAGRLI 512  
 |||||  
 Db 1800 DGRLLKAGRLI 1810

## RESULT 4

S76086  
 beta transducin-like protein, 190K - Synechocystis sp. (strain PCC 6803)  
 N:Alternate names: protein sl10163  
 C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence, revision 25-Apr-1997 #text\_change 26-May-2000  
 R:Accession: S76086  
 R:Kneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 O., K.; Okunura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, S.  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC 6803  
 A:Reference number: S74322; MUID:97061201

A:Accession: S76086  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1693 <KAN>  
 A:Cross-references: EMBL:DB3999; GB:AB001339; NID:q1001396; PIDN:BA10064.1; PID:d101  
 F:1051-1084/Domain: WD repeat homology <WD01>  
 F:1092-1125/Domain: WD repeat homology <WD02>  
 F:1133-1166/Domain: WD repeat homology <WD03>  
 F:1174-1207/Domain: WD repeat homology <WD04>  
 F:1256-1289/Domain: WD repeat homology <WD05>  
 F:1297-1330/Domain: WD repeat homology <WD06>  
 F:1338-1371/Domain: WD repeat homology <WD07>  
 F:1420-1453/Domain: WD repeat homology <WD08>  
 F:1461-1494/Domain: WD repeat homology <WD09>  
 F:1502-1535/Domain: WD repeat homology <WD10>  
 F:1584-1617/Domain: WD repeat homology <WD11>  
 F:1625-1658/Domain: WD repeat homology <WD12>

Query Match 0.7%; Score 9; DB 2; Length 1693;  
 Best Local Similarity 100.0%; Pred. No. 8.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 381 RPSPSQYI 389  
 |||||  
 Db 1431 RPSPSQYI 1439

## RESULT 5

T15027  
 hypothetical protein Y1112 - Yersinia pestis plasmid pMT1  
 C:Species: Yersinia pestis  
 C:Date: 20-Sep-1999 #sequence, revision 20-Sep-1999 #text\_change 20-Sep-1999  
 R:Accession: T15027  
 R:Lindler, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.  
 Infect. Immun. 66, 5731-5742, 1998  
 A:Title: Complete DNA sequence and detailed analysis of the Yersinia pestis KIM5 plasmid  
 A:Reference number: Z18268; MUID:99043898  
 A:Accession: T15027  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-88 <LIN>  
 A:Cross-references: EMBL:AF074611; NID:93883003; PID:93883110; PIDN:AAC2770.1  
 C:Genetics:  
 A:Gene: Y1112  
 A:Genome: plasmid pMT1

Query Match 0.6%; Score 8; DB 2; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 6.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 622 KSIINGCA 629  
 |||||  
 Db 3 KSIINGCA 10

## RESULT 6

T39293  
 hypothetical protein SPB1105.16c - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence, revision 03-Dec-1999 #text\_change 03-Dec-1999  
 R:Accession: T39293  
 R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Ransperger, U.; Bothe, G.; Pohl, A.  
 submitted to the EMBL Data Library, July 1999  
 A:Reference number: Z21841  
 A:Accession: T39293  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-107 <MCD>  
 A:Cross-references: EMBL:AL096851; PIDN:CAB50979.1; GSPDB:GN00067; SPDB:SPB1105.16c



A:Experimental source: strain 972h-; cosmid c1105  
 C:Genetics:  
 A:Gene: SPDB:SPC1105.16c  
 A:Map position: 2

Query Match 0.6%; Score 8; DB 2; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 7.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 EEPKRNK 170  
 DB 77 EEPKRNK 84

## RESULT 7

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: C86179

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maitl, R.; Marziani, Rizzo, M.; Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719

A:Accession: C86179

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-128 <STO>

A:Cross-references: GB:AE005172; NID:g2494117; PIDN:AAB80626.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 0.6%; Score 8; DB 2; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 284 LGIHVPF 291  
 DB 53 LGIHVPF 60

## RESULT 8

probable cation-activated repressor protein (tror) - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 28-Jul-2000

C:Accession: D71357

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwaltson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utechtack, T.; McDev, they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770

A:Accession: D71357

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-153 <COL>

A:Cross-references: GB:AE001201; GB:AE000520; NID:g3322431; PIDN:AAC65157.1; PID:g332243

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0167

C:Superfamily: conserved hypothetical protein M0568

Query Match 0.6%; Score 8; DB 1; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 303 VKRLEKGG 310  
 DB 48 VKRLEKGG 55

## RESULT 9

hypothetical protein BH3741 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: E84117

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: E84117

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-219 <STO>

A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA07460.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3741

Query Match 0.6%; Score 8; DB 2; Length 219;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 507 AGDRLIEV 514  
 DB 41 AGDRLIEV 48

## RESULT 10

hypothetical protein M0635 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000

C:Accession: C64379

R:Balt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak R:Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc

A:Reference number: A64300; MUID:96337999

A:Accession: C64379

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-283 <BU>

A:Cross-references: GB:U67511; GB:L77117; NID:g28263300; PIDN:AAB98633.1; PID:g1591347

C:Genetics:

A:Map position: E0R564790-565641

A:Start codon: GTC

Query Match 0.6%; Score 8; DB 2; Length 283;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 999 KKKDRDKE 1006  
 DB 245 KKKDRDKE 252

## RESULT 11

hypothetical protein Y54G11A.10 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C;Accession: T27179  
 R;Walls, J.  
 Submitted to the EMBL Data Library, December 1998  
 A;Reference number: Z20322  
 A;Accession: T27179  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-317 <WIL>  
 A;Cross-references: EMBL:AL034488; NID:e1359895; PIDN:CAA22459.1; CESP:Y54G11A.10  
 A;Experimental source: clone Y54G11A  
 C;Genetics:  
 A;Gene: CESP:Y54G11A.10  
 A;introns: 55/3; 108/3; 175/3; 228/2; 253/3

Query Match 0.6%; Score 8; DB 2; Length 317;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 640 DDLIANG 647  
 |||||||  
 Db 245 DDLIANG 252

RESULT 12  
 T24509  
 hypothetical protein T05C12.8 - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C;Accession: T24509  
 R;Burton, J.  
 Submitted to the EMBL Data Library, October 1995  
 A;Reference number: Z19901  
 A;Accession: T24509  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-318 <WIL>  
 A;Cross-references: EMBL:Z65500; PIDN:CAA91309.1; GSPDB:GN00020; CESP:T05C12.8  
 A;Experimental source: clone T05C12  
 C;Genetics:  
 A;Gene: CESP:T05C12.8  
 A;Map position: 2  
 A;introns: 26/3; 65/3; 103/3; 204/3; 245/3

Query Match 0.6%; Score 8; DB 2; Length 318;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1302 PSEGPSN 1309  
 |||||||  
 Db 261 PSEGPSN 268

RESULT 13  
 149262  
 cyclin cdk inhibitor p57 - mouse  
 N;Alternate names: CDI p57; cyclin-cyclin-dependent kinase inhibitor p57; p57-kip2  
 C;Species: Mus musculus (house mouse)  
 C;Date: 09-Mar-1996 #sequence\_revision 09-Mar-1996 #text\_change 05-Nov-1999  
 C;Accession: 149262  
 R;Lee, M.H.; Reynolds, I.; Massague, J.  
 Genes Dev. 9, 639-649, 1995  
 A;Title: Cloning of p57KIP2, a cyclin-dependent kinase inhibitor with unique domain structure  
 A;Reference number: A56463; MUID:95247027  
 A;Accession: 149262  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-348 <RES>  
 A;Cross-references: EMBL:U20553; NID:g794295; PIDN:AAC52186.1; PID:g794296  
 C;Keywords: cell cycle control

Query Match 0.6%; Score 8; DB 2; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 431 PPSAPASA 438  
 |||||||  
 Db 139 PPSAPASA 146

RESULT 14  
 AD0380  
 multidrug efflux protein [imported] - Yersinia pestis (strain C092)  
 C;Species: Yersinia pestis  
 C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
 C;Accession: AD0380  
 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tittball, R.W.; Holden, M.T.G.; Prentice, M.  
 deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.  
 Nature 413, 523-527, 2001  
 A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A;Reference number: AB0001; MUID:21470413; PMID:11586360  
 A;Accession: AD0380  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-395 <KUR>  
 A;Cross-references: GB:AL590842; PIDN:CAC92367.1; PID:g15981070; GSPDB:GN00175  
 C;Genetics:  
 A;Gene: acra

Query Match 0.6%; Score 8; DB 2; Length 395;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 505 LKAGDRLI 512  
 |||||||  
 Db 354 LKAGDRLI 361

RESULT 15  
 G01158  
 tyrosine kinase activator protein 1 (TKA-1) - human  
 C;Species: Homo sapiens (man)  
 C;Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 16-Dec-1998  
 C;Accession: G01158  
 R;Closek, T.  
 Submitted to the EMBL Data Library, July 1995  
 A;Reference number: H00587  
 A;Accession: G01158  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-450 <CTO>  
 A;Cross-references: EMBL:Z50150; NID:g1246762  
 C;Superfamily: GIGF domain homology  
 F;16-86/Domain: GIGF domain homology <GLG1>

Query Match 0.6%; Score 8; DB 2; Length 450;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 509 DRLIEVNG 516  
 |||||||  
 Db 193 DRLIEVNG 200

Search completed: July 24, 2002, 14:36:07  
 Job time: 158 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 06:40:28 ; Search time 6663.82/Seconds

(without alignments)  
17865.288 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689  
Sequence: 1 atgaagtgaccgtgtcctt.....tcttcctgttaaaaaaaaaa 5689

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_da:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mn:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match Length	DB ID	Description
------------	-------	-------	--------------	-------	-------------

1	5636.8	99.1	5958	9	AF196185	AF196185 Homo sapi
2	4069.4	71.5	4071	9	AF467002	AF467002 Homo sapi
3	4041.4	71.0	4062	9	AF467003	AF467003 Homo sapi
4	3837.4	67.5	3960	9	AF467004	AF467004 Homo sapi
5	3763.4	66.2	3933	9	AF467006	AF467006 Homo sapi
6	3529.8	62.0	3822	9	AF332593	AF332593 Homo sapi
7	3486.2	61.3	3801	9	AF252293	AF252293 Homo sapi
8	3347.4	58.8	3735	9	AF467005	AF467005 Homo sapi
9	3056.4	53.7	5500	10	AB005549	AB005549 Rattus no
10	3050.6	53.6	3470	9	AF196186	AF196186 Homo sapi
11	2952.2	51.9	4167	10	AY026057	AY026057 Mus muscu
12	2772.6	48.7	2967	9	AF332592	AF332592 Homo sapi
13	2677.4	47.1	2696	9	AK000761	AK000761 Homo sapi
14	2476.8	43.5	2718	6	BD004905	BD004905 Novel gen
15	2476.8	43.5	2718	9	AK027735	AK027735 Homo sapi
16	2079.4	36.6	2116	9	AK024668	AK024668 Homo sapi
17	2005.2	35.2	196374	2	AL591464	AL591464 Homo sapi
18	2003.6	35.2	106198	9	AL160409	AL160409 Human DNA
19	2003.6	33.2	138688	2	AC025823	AC025823 Homo sapi
20	1982.2	34.8	22279	2	U82210	U82210 Homo sapien
21	1868.8	32.8	2004	9	AF177228	AF177228 Homo sapi
22	1695	29.8	1772	9	AK025892	AK025892 Homo sapi
23	1546.8	27.2	1592	9	BC011711	BC011711 Homo sapi
24	516.2	9.1	597	6	AX187060	AX187060 Sequence
25	482.4	8.5	540	6	AX321805	AX321805 Sequence
26	427	7.5	427	6	AX332939	AX332939 Sequence
27	415.4	7.3	425	6	AX072873	AX072873 Sequence
28	387	6.8	104357	9	AL392123	AL392123 Human DNA
29	357.4	6.3	504	6	AX185634	AX185634 Sequence
30	282.6	5.0	305	6	AX187694	AX187694 Sequence
31	282.6	5.0	309	6	AX186627	AX186627 Sequence
32	282.6	5.0	168213	2	AC103175	AC103175 Rattus no
33	273.2	4.8	4176	9	AF428250	AF428250 Homo sapi
34	271.6	4.8	339	6	AX184822	AX184822 Sequence
35	261.6	4.5	2656	9	AK056157	AK056157 Homo sapi
36	256.8	4.5	2068	9	AK057965	AK057965 Homo sapi
37	253.8	4.5	400	6	AX071341	AX071341 Sequence
38	246.8	4.3	70087	2	AC105981	AC105981 Mus muscu
39	245.4	4.3	165146	2	AC011644	AC011644 Homo sapi
40	245.4	4.3	167163	2	AC013705	AC013705 Homo sapi
41	245.4	4.3	169337	2	AL356462	AL356462 Homo sapi
42	245.4	4.3	193131	9	AL390766	AL390766 Human DNA
43	182.8	3.2	109816	9	AL360233	AL360233 Human DNA
44	182.8	3.2	138322	2	AC012241	AC012241 Homo sapi
45	182.8	3.2	196259	2	AL392171	AL392171 Homo sapi

## ALIGNMENTS

RESULT	1	AF196185	5958 bp	mRNA	linear	PRI 16-NOV-2001
LOCUS	AF196185					
DEFINITION	Homo sapiens atypical PKC isotype-specific interacting protein long					
ACCESSION	AF196185					
VERSION	AF196185.1					GI:13491609
KEYWORDS						
SOURCE						
ORGANISM	human.					
REFERENCE						
AUTHORS	Fukuyota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.					
JOURNAL	Fang, C.M. and Xu, Y.H.					
PUBMED	Down-regulated expression of atypical PKC-binding domain deleted					
AUTHORS	Cell Res. 11 (3), 223-229 (2001)					
TITLE	Cell Res. 11 (3), 223-229 (2001)					
POLARITY	Gene, HsPip					

DATE - SEP 7

Unpublished  
3 (bases 1 to 5958)  
Fang, C. and Xu, Y.  
Direct Submission  
Submitted (18-Oct-19)

Oncology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-yue Road, Shanghai 200031, China

**Source**

Association/Qualifiers  
1. .5958  
Shanghai 200031, P.R. China

**CDS**

```

/alias="sarsin" homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
288..4349
/note="HAS1PIA"
/codon_start=1
/product="atypical PKC isotype-specific interacting
protein long variant"
/protein_id="AAK27891.1"
/db_xref="GI:13491610"

```

BASE COUNT	1623 a	1395 c	1581 g	1358 t	1 045
ORIGIN	PAAAGPRDVPSPSQVARKLRRLQTPDKGRPFYS"				

Query Match	99.1%	Score 5636.8	DB 9;	Length 5958;
Best Local Similarity	99.8%;			
Matches 5668;	Conservation			

	1	atgaagtgcacgtatgcttcgaagcaaa	Indels	10;	Gaps	2
--	---	------------------------------	--------	-----	------	---

288 ATGAAAGTGACCGTGTCTTCGGACGACCCCGGGTGGTTGCACCACTCCTT

348  
123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899100101102103104105106107108109110111112113114115116117118119120121122123124125126127128129130131132133134135136137138139140141142143144145146147148149150151152153154155156157158159160161162163164165166167168169170171172173174175176177178179180181182183184185186187188189190191192193194195196197198199200201202203204205206207208209210211212213214215216217218219220221222223224225226227228229230231232233234235236237238239240241242243244245246247248249250251252253254255256257258259260261262263264265266267268269270271272273274275276277278279280281282283284285286287288289290291292293294295296297298299300301302303304305306307308309310311312313314315316317318319320321322323324325326327328329330331332333334335336337338339340341342343344345346347348349350351352353354355356357358359360361362363364365366367368369370371372373374375376377378379380381382383384385386387388389390391392393394395396397398399400401402403404405406407408409410411412413414415416417418419420421422423424425426427428429430431432433434435436437438439440441442443444445446447448449450451452453454455456457458459460461462463464465466467468469470471472473474475476477478479480481482483484485486487488489490491492493494495496497498499500501502503504505506507508509510511512513514515516517518519520521522523524525526527528529530531532533534535536537538539540541542543544545546547548549550551552553554555556557558559560561562563564565566567568569570571572573574575576577578579580581582583584585586587588589590591592593594595596597598599600601602603604605606607608609610611612613614615616617618619620621622623624625626627628629630631632633634635636637638639640641642643644645646647648649650651652653654655656657658659660661662663664665666667668669670671672673674675676677678679680681682683684685686687688689690691692693694695696697698699700701702703704705706707708709710711712713714715716717718719720721722723724725726727728729730731732733734735736737738739740741742743744745746747748749750751752753754755756757758759760761762763764765766767768769770771772773774775776777778779780781782783784785786787788789790791792793794795796797798799800801802803804805806807808809810811812813814815816817818819820821822823824825826827828829830831832833834835836837838839840841842843844845846847848849850851852853854855856857858859860861862863864865866867868869870871872873874875876877878879880881882883884885886887888889890891892893894895896897898899900901902903904905906907908909910911912913914915916917918919920921922923924925926927928929930931932933934935936937938939940941942943944945946947948949950951952953954955956957958959960961962963964965966967968969970971972973974975976977978979980981982983984985986987988989990991992993994995996997998999100010011002100310041005100610071008100910101011101210131014101510161017101810191020102110221023102410251026102710281029103010311032103310341035103610371038103910401041104210431044104510461047104810491050105110521053105410551056105710581059106010611062106310641065106610671068106910701071107210731074107510761077107810791080108110821083108410851086108710881089109010911092109310941095109610971098109911001101110211031104110511061107110811091110111111121113111411151116111711181119112011211122112311241125112611271128112911301131113211331134113511361137113811391140114111421143114411451146114711481149115011511152115311541155115611571158115911601161116211631164116511661167116811691170117111721173117411751176117711781179118011811182118311841185118611871188118911901191119211931194119511961197119811991200120112021203120412051206120712081209121012111212121312141215121612171218121912201221122212231224122512261227122812291230123112321233123412351236123712381239124012411242124312441245124612471248124912501251125212531254125512561257125812591260126112621263126412651266126712681269127012711272127312741275127612771278127912801281128212831284128512861287128812891290129112921293129412951296129712981299

121 gatcgaactatccct-  
-----gcacgucgttgaccgccgttacggagsgcccatgcccaag 40

[illegible]

181 ctgtagacattcttgtgtatgttagcagacgataaaagacagactctt

341  
-----GACCAICTTTGTGATGTAGCAGACGATAAGACAGACTGGTAGCAGTGTTCAT 533

528 GAGCAGGATGCGGACCAAGTGCAGTCCACGCGTACCAAGAGC 300

301 ccagagatatlttgtaqtaaaccttcccaacacccagc 587

360

[illegible]

Dh 1668 AATATCCAGCTTAAAGAACGTACAGAACGTTGGATTGACATACCTTCCAGACATGTA 1727  
Qy 1441 acaataggctgctcagctcccaatctatg tgaataacatctcccccgggggagccat 1500  
Dh 1728 ACAATAGGTGGCTGACGTCCATCTATGTGAATAAATCTTCCCGGGGGGGGCGCANT 1787  
Qy 1501 cagatgagcgaacttaagcagagacagacttaagatgaatgagtagattagtg 1560  
Dh 1788 CAGATGGCCCACTTAAAGCAGAGACACTTATAGAGTAAATGAGTAAATTTAGT 1847  
Qy 1561 ggaacatcccaaggagagattgcttcgctgtgtgagaagacccaagaatgagaagacttg 1620  
Dh 1848 GCGAATATCCAGAGAGAGATTGTTCTGCTGTGAGAACACCAAGATGGAGAGAACTGG 1907  
Qy 1621 agccttcgtgcttcctcgcaggaagacgcttcacccaagggaactgaatgacagca 1680  
Dh 1908 AGCCTTGTGCTTTCGCCAGGAAGACGCTTCCACCCAGGAAGCTGATGACAGCA 1967  
Qy 1681 agcgaatgacgaatccaaaagaagaagaagaatgagaatattgcttcaacct 1740  
Dh 1968 AGCCAGATGAGATTCCAAAAGAAACGAGAAAGATGAGATATTGTTCTACACT 2027  
Qy 1741 gatgacacaaaggaaattctacattgaagtcaccaataatgattcaggatctcagc 1800  
Dh 2028 GATGGCACCGAGGAAATTTCTGACATTGAAGTCCACTTAATGATTCAGGATCTGAGGC 2087  
Qy 1801 ctgtgtgctcagtgctcaaaagtaaccggtcaaaagaacacagcagatttggaaactt 1860  
Dh 2088 CTTGGTGTCACTGTCAAAAGTAACCGGTCAAAAGAAACACGAGATTGGGAATCTT 2147  
Qy 1861 gtcaagtcattatgaatgagagagacatctaaagatggaagcttcgggtgaatg 1920  
Dh 2148 GTCAAGTCCATTATTAAGGAGAGACACATCAAAAGATGAAAGCTTCGGGTGATGAT 2207  
Qy 1921 caactgatagagtaaatggaggaatccctgtgtggcaagacaacaaagatgccaatgaa 1980  
Dh 2208 CAACTGATAGCAATTAATGAGAAATCCCTGTGTGGCAAGCAACCAAGATGCCATGGA 2267  
Qy 1981 accctaagaaggtctatgtctactgaagcaataaagcaggaatgatccagcttattgt 2040  
Dh 2268 ACCCTAAGAGGTCTATGTCTACTGTAAGCAATTAACGAGGAATATCCAGTTATTGT 2327  
Qy 2041 gcaagaggaataaagaagtgcaatgagctgaagtaacctgaggagccccctgagacctag 2100  
Dh 2328 GCAAGGAATAATAGCAATGTGAAATGAGTGAAGTCACTGGAGGCCCCCTGAGACTGAG 2387  
Qy 2101 ctgcccattgaaacagcgttgatgataagagaacgaagaatttccattccctctacagt 2160  
Dh 2388 CTGCCCATTTGAACAGCTTGGATGATAGAGAACGAAAGATTTCCCATTTCCCTACAGT 2447  
Qy 2161 ggaattgaggggcttgatgaatcgccccagcagaatgctgcctcagtagaataatgggt 2220  
Dh 2448 GGGATTGAGGGGCTTGATGAATCGCCAGCAGAATGCTGCCCTCAGTAGAGTAATG--- 2504  
Qy 2221 gagtcaggttaataaccagctgtccctctacagtgaataatgccccagaatgacagcttcat 2280  
Dh 2505 -----GGTAAATACACACTGCCCCCTACAGTGAATATGCCCAAGATGACACTGTCTAT 2558  
Qy 2281 atagaagaatgacaggttgccagtgcttcctcacatctcttgaccagcttccctccagc 2340  
Dh 2559 ATAGAAGATGACAGGTTCCAGTGCTTCTCCACATCTCTCTGACACAGTCCCTTTCCAGC 2618  
Qy 2341 tcccaatgataatgaggttctgtagcggcagatgctgg laacttggccaaggtcgaatc 2400  
Dh 2619 TCCCATGATGATGTGGGTTGTGTGACGGCAGATGCTGGTACTTGGCCCAAGCTCAGATC 2678  
Qy 2401 agtgaattcagcagctgtctcttgagctccagatgctgacagcttctgcttccaaga 2460  
Dh 2679 AGTGAATTAGCCGACTGCTCTTGTGAGTCCAGATGTTGATCCAGTCTTCTTTTAAACA 2738  
Qy 2461 gaaggaatttgacgtcagagatgctcagaaaaagcacaaagcaatttcagatgacagt 2520  
Dh 2739 GAAGGATTTGACGTCAGAGTATGTACAAAAACGCAAAAGCAATTTTACATGCGCAGT 2798  
Qy 2521 caattggattctgttaaaaaacagaaatcaaaaaagcatgattagtaatactgacag 2580  
Dh 2799 CAATTGATTTGCTTAATAAACACGAAATCAAAAAGCATGATTTAGTATGAGTGACGAG 2858  
Qy 2581 actaaactcaatacagtgatgatacgaagaagaagcttctccagcagagatgtggctct 2640  
Dh 2859 ACTAAACTCAATACAGTGTGATGACCGAAGAGCAGGTTCTCCACGACGAGATGTGGGTCT 2918  
Qy 2641 tccctgggtctgaagaagtcagagctcggtggagagctctcagaacgcaatttcccgagtg 2700  
Dh 2919 TCCCTGGGTCTGAAGAAGTCAAGCTCATTGAGAGTCTCAGACGCCCACTTGCAGAGTG 2978  
Qy 2701 acttgaatgaggataatctcttccatcgtccagggccggcggataatcagagcagggga 2760  
Dh 2979 ACTTGAATGGGGAATTTCTTCCATCTCCACGGCCCGGATATACAGAGCAGAGGGA 3038  
Qy 2761 tgcgaatgagagcttcagagctgcacatcgacaatcttatgataaaccgcytatgat 2820  
Dh 3039 TGCATGAGAGCTTCAGACCTGCCATCGACAAATCTTATGATTAACCCCGGTAGATGAT 3098  
Qy 2821 gatgatgaaggcatggaagccttggaagaagacacagaagaagatccaagatccaaggaga 2880  
Dh 3099 GATGATGAAGGCAATGGAGACTTGTGAGAGACACAGAAAGAAATTCAGATCAGGAGA 3158  
Qy 2881 gagctgtatccacagccagtgatcagccttccactctctgtgagagacaaatgaaatgga 2940  
Dh 3159 GAGTGTGATCCACAGCCAGTGTACGCTTCCACTCTGTGAGAGACAAATGATGGA 3218  
Qy 2941 aaccaagagaagtgatgataagctgataaaaaaagaataaaactggaagaagaagaag 3000  
Dh 3219 AACCAAGAGAAAGGTGATTAAGACTGATGAAAAAAGATTAAGATGAAAAAGAAAG 3278  
Qy 3001 aagatlaagatlaagagaagaataaaatgaaagccaagaaggaatgcttgaaggcttg 3060  
Dh 3279 AAGATAGAGATAGAGAGAGAGATTAATTAATGAAGCCAAAGAGGGAATGCTGAAGGCTTG 3338  
Qy 3061 ggaagacatgtcaaggttggcaaacatcgaaaagatgaaagaatggaagaacggtlaaa 3120  
Dh 3339 GGAGACATGTTCAGGTTTGGCAAAACATCGMAAAGATGACAAAGATTGAAAAACGGGTAA 3398  
Qy 3121 ataaaaatcagaagatccttatacagaagaaggagagatacgaatgaagcagaagcag 3180  
Dh 3399 ATTAATAATACAGAAATCCCTTACATCAGAAAGAGACAGATACGAATGAACGGGAGAG 3458  
Qy 3181 gagagattcagaaccaaactcgagaatlaaggaacagcagctgagaagcgtgactat 3240  
Dh 3459 GAGAGATTCAAGCCCAAACTCGAGAAATTTAGGGAACGACAGCTCGAGGCTGACTAT 3518  
Qy 3241 gctgaattcagaatttcatcaggaatttggctgtagatgataatgtaatgtaatgga 3300  
Dh 3519 GCTGAATTAAGATTTTCATCGAGATTTGGCTGTGATGATGATGATGATGATGATG 3578  
Qy 3301 gttcttcttaagaaggttcacatgctcacaagctagaacctcagaagcccaagaaggg 3360  
Dh 3579 GTTTCTTCTTAAGAAAGGTTCATGCTCAACGCTAGACCTCAGAGCCACAGAAAGG 3638  
Qy 3361 catatgatgattgcttgtatgcccagtcagaagaacggcggaattccaaacccctaacct 3420  
Dh 3639 CATATGATGATGCTTTTGTATGCCCCAAGTCAMAGAACCCCGGAATTCCAAACCTTACCT 3698  
Qy 3421 gtagaacgtaacagatcaactcctcagaatcatgatacgaatacagcgttgaagcaaga 3480  
Dh 3699 GTAGACAGTAAACAGATCAACTCTACATCATGATCGGATACAGGCTGAGGCAAGAA 3758  
Qy 3481 ttccagcaagcaagaagaatgaagaatgataagaatgctcggcggactatagtttgag 3540  
Dh 3759 TTTTCAGCAAGCAAGCAAGATGAAGATGAAGATGCTCGCGGACCTATAGTTTGAAG 3818  
Qy 3541 caaaccttggcgaacagcagcggcgagcagagagcggcgcaactcgggtgtccgtggag 3600  
Dh 3819 CAACCTGGCCGAAACGACGCGCGGCGAGACGACGCGGCGACACTCGGTTCCGTGAG 3878

[illegible][illegible]

RESULT	2
AF467002	
LOCUS	
DEFINITION	AF467002 Homo sapiens partitioning-defective 3 protein splice variant a (PARD3) mRNA, complete cds, alternatively spliced.

ACCESSION AF467002  
VERSION AF467002.1 GI:18568343  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 4071)  
AUTHORS Gao, L., Macara, I.G. and Joberty, G.  
TITLE Multiple splice variants of PAR3 and of a novel related gene,  
PAR3L, produce functionally different proteins  
JOURNAL 2 (bases 1 to 4071)  
REFERENCE Gao, L., Macara, I.G. and Joberty, G.  
AUTHORS Direct Submission  
JOURNAL Submitted (10-JAN-2002) Center for Cell Signaling, University of  
Virginia, PO Box 800577, Charlottesville, VA 22908, USA  
FEATURES  
SOURCE  
1..4071  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..4071  
/gene="PAR3"  
/note="PAR3"  
1..4071  
/gene="PAR3"  
/note="PAR3; multi-PDZ domain containing protein;  
alternatively spliced"  
/codon\_start=1  
/product="partitioning-defective 3 protein splice variant  
a"  
/protein\_id="AA176042.1"  
/db\_xref="GI:18568344"  
/translation="MKTVVCFGRVVVPCGDGHMKVSLIQAVTRYKAIADPNY  
WIOYRLHGGGGGLDDLDLIDCDVADDERDRIYAVEDEDPHGGGTSASSGTGSP  
IFGSEIGTNNYSAROPYOATSEIVTPEVLAMNDELHYRSDPALIGLSTSVSNF  
SSRPSKRNPTRMSTTAAFLKQNTAGSKTCDKRDENRSLPRTSMWSNQFDNA  
RSLLSAPHVGMLEKQDEDEGTEEDNSVEPYGHADTGLEHLPNSLDMVRLVE  
VPNDGGPGIHVFPSSARGRTGLVLRLEKGGAEHEHNFRENDCLVRINDGLNR  
RNFEOAOHFRQAMKPTPIIMFHVVPANKEQYEQLSQEKNNYSRSFSPQYIDNR  
SNSAGLHTVORAPRLNHPPEIDSHSLPSPHSAHSGKPPSPAPAPQNVSTVSG  
YNTKIGKRLNIOLKKEGEGISITSDVITIGSAPLYVNIILPRGAIDDLKAG  
DLLELVNGVDLVKSGQEVVSLSTRKEGVSLVFNQEDAFHRELNAESQKQIP  
KETAEDEDIVLPDGTRETFEYVPLNDSGAGVSKGRNSKENADIGIPKSI  
INGKASDKGLRVNDOLIAVNGESLIGKTQDAETRLRSMTEGNRGMIOILVAR  
RISGCNELKSGSPPELPJETALDDEKRIISHSLYSGIIGLDESPRNALSTRIMG  
ESGYQLSPYNNMPQDDVITIEDRLPYLPPLHSSSSSSHDGVTADAGTWAKA  
AISDSADCSLSPVDVYLAFOREGRGSMSEKRTKOPSDASOLDFVTRKSKMDIG  
IADPRKLTNDVDOKAGSPRDVPSLGLKSSLSLEIOTNAEVTLLNDIPRHRPR  
IIRGCGNESFRAIIDSYDKPAVDDDEGMEULEEDTEESRSRESVSTASDP  
SLERQMNQNGDKTDRKDKTGKKEKDKEDKDKKAKGMLKGLDMERFGKHR  
KDKRIETKGIKIGESFSEERIIMKOEORIKQKTEFERROREDVYEIOPFHR  
TFCCDELMYGVSSYEGSMALNARPOSGREGHMDALYAQVKRKNKSPSVDNRS  
TFSNHDRIORLROEFOQKODEVEDRRTATSPEDPNARAPATGSHKSPVEMO  
RROEERESSQOAROYSLPROSKKNSYSQSWKQNPSPGEPFOGAKKEPSSY  
QGSRNGYLGGCFNARVYLETQELRLDORRKEOQMKQPSSEGSNTDSTKRVDDPS  
YAPKGPEDVPPSPSOVARLNRLQTEKRPFS"  
814..1074  
/gene="PAR3"  
/note="Region: PDZ domain"  
1378..1638  
/gene="PAR3"  
/note="Region: PDZ domain"  
1765..2028  
/gene="PAR3"  
/note="Region: PDZ domain"  
BASE COUNT 1220 a 923 c 1079 g 849 t  
ORIGIN

Query Match 71.5%; Score 4069.4; DB 9; Length 4071;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgaagtgaccgtgtcttcggaacgaccggtgtgtctgcgttgcgggagcgccac 60  
|||||  
Db 1 ATGAAGTGACCGTGTCTTCGGACGACCGGGGTGTCTGCCGTGGCGGACGGCCAC 60  
QY 61 atgaagtttcaagcctcatccagcagcggtgtgacccgtaccggaaagcgccaa 120  
|||||  
Db 61 ATGAAGTTTCAAGCCTCATCCAGCAGCGGGTGTACCCGCTACCGGAAGCGCATCGCCAA 120  
QY 121 gatccaactactggtatcagtgatcgtcttggaacatgagatgagaaactagac 180  
|||||  
Db 121 GATCCAACACTACTGATACAGTGATCGCTTGGAACATGGAGATGAGGAATACTAGAC 180  
QY 181 ctgtatgatacttctgtatgtagaagacgataaagacagctgtgtagcagtttgat 240  
|||||  
Db 181 CTGTATGATACATCTCTGTGATGTAGCAGACGATAAAGACAGACTGTGAGCAGTGTGAT 240  
QY 241 gaggcagatcccaatacgaagtgatgtagcagcagtgccaggtccacggtaccagagc 300  
|||||  
Db 241 GAGCAGATCCCATATCAGGAGGTGATGGCACCAGTGCCAGTCCACGGGTACCCAGAGC 300  
QY 301 ccagagatatgttagtagtgcgttggaaccaaatgtctcagccttcaagccttacc 360  
|||||  
Db 301 CCAGAGATATTTGTGTAGTGAAGCTTGGCACCACCAATGTCTCAGCCTTTACACCA 360  
QY 361 gcaacaagtgaatltgagtgacacactcaagtccttcgagaaatatycccttcatt 420  
|||||  
Db 361 GCACAAGTGAATGTAGGTGACACACTTCACTGCTTCGACCAATATATGCTTTCATGTT 420  
QY 421 cgaagcagtagtagtaccacagctctcaattgagccttccactctgtcagtgtagtaattt 480  
|||||  
Db 421 CGAGCAGTAGTAGTACCACAGCTCTCAATTTGGCTTCCACATTTCTGAGTGAATTAATTTT 480  
QY 481 tccctcgaagagccttcaagaaataaccacagcgtgtgtcaacaacagctgtcttc 540  
|||||  
Db 481 TCCTCTGAAGAGCCTTCAAGGAATAATCCACACAGCTGTGCAACACAGCTGTCTCTC 540  
QY 541 aagcagaacaactgtcgtggaagtcctaaacccctgcgacagaagaagaatgaatacaga 600  
|||||  
Db 541 AAGCAACAACACTCTGTGGAGTCTTAAACCTCGACAGNAAGNAAGTGAATACTACAGA 600  
QY 601 agcctccgcggagatctctgtagatctgtctcaacaaatttaagagagaagaatgtctgc 660  
|||||  
Db 601 AGCCTCCCGGGATCTGTAAGTCTGTCAACCAATTTCTAGAGAGCAATATGCTCTCCTGC 660  
QY 661 tctctagtcacagtcacccaatgtgtggaagtcgctgtagaagaacaagaacagagatga 720  
|||||  
Db 661 TCTCTAGTGCACGTACCCAAATGTGTGGCAATGTGCTGAGAAACAAGACAGATGAG 720  
QY 721 gatggacaagaagataacagtcgtgtgaaacctgttgacacgtctgaacagggattg 780  
|||||  
Db 721 GATGGACAGAAAGATTAACAGTCTGTGAACCTGTTGACATGCTTGACACGGGTTTG 780  
QY 781 ggcacatacccaacttctctcgtgtagatgtgtagaagtcgtagaagcccccaagat 840  
|||||  
Db 781 GACACATATACCCCACTTTCTCTGAGATATGTGTAAGCTCTAGAAAGTCCCAACGAT 840  
QY 841 ggaaggcctctgggaatccaatgagtcagtccttcaagtcgtcgaagcgcaaaccttgagg 900  
|||||  
Db 841 GAGAGGCTCTCTGGGAATCCATGATAGTGTGAGTGTGAGAGCGGCAACACCTGTGGG 900  
QY 901 ttaattagtaaacgatttgaagaagtggtgaagcgtgaacatgaataacttcttgtag 960  
|||||  
Db 901 TTATTAGTAAACGATTGTGGAAGAGTGTGAAGCTGAAACATGAAATCTTTTCCGTGAG 960  
QY 961 aatgattgcatgtgtcaggttaatgtagtgagaccttcgaataatgaagatttgaacaga 1020  
|||||  
Db 961 AATGATTGCAATGTGTGAGATTAATGATGGGACCTTTCGAATTAAGAAATTTGAACAAGCA 1020  
QY 1021 caacataatgttcgcaagccatgctgtaacccatatttgttccatgtgtgtctcgca 1080  
|||||  
Db 1021 CAACATATGTTCGCCAAGCGATGCTTACACCATCATATTGTGTTCCATGTGTCTCTGCA 1080

QY 1081 gcaataaagacagatgatgaacactatcccaagtgaagaagaacatactatccaagc 1140  
 Db 1081 GCAAAATAAAGAGAGATGATGAACACATATCCCAAGTGAGAAACAAATATATTCAAAGC 1140  
 QY 1141 cgttttagccttgacagccagatatttgacaacagaagtgtgaacagtgacagccttcac 1200  
 Db 1141 CGTTTACCCCTGACACCCAGTATATTGCAACACAGAGCTGTGAACACAGTGCAGGGCTTCAC 1200  
 QY 1201 acagtgacagacacccagcagatgaaccccgccctgaacagatagactctacataaga 1260  
 Db 1201 ACAGTGACAGACACCCAGCTGAAACCAACCCGCTGAGAGATAGACTCACTCACTCAAGA 1260  
 QY 1261 ctacccatagacgacacccctcggaacacacacacacacacacacacacacacacacac 1260  
 Db 1261 CTACCCATAGACGACACCCCTCGGGAACCAACACACACACACACACACACACACACAC 1260  
 QY 1321 aatgatattagtagactgtgaacagtgatgaacacacacacacacacacacacacacacac 1320  
 Db 1321 AATGATATTAGTAGACTGTGAACAGTGTGATGAACACACACACACACACACACACACAC 1320  
 QY 1381 aatagcagcttaagaaagtlacagaagttgggattcagcactcagcactcagcactcagc 1380  
 Db 1381 AATAGCAGCTTAAGAAAGTTLACAGAAGTTGGGATTTCAGCACTCAGCACTCAGCACTCAG 1380  
 QY 1441 acatagtgagcagcagcactcactcactcactcactcactcactcactcactcactcactc 1440  
 Db 1441 ACAATAGTGAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTC 1440  
 QY 1501 caagatgacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1500  
 Db 1501 CAGGATGCGGCGAGCTTAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 QY 1561 gacaaatcccaag 1560  
 Db 1561 GACAAATCCCAAG 1560  
 QY 1621 agccttcggtccttcgacag 1620  
 Db 1621 AGCCTTGCTGCTTTCGCGAG 1620  
 QY 1681 agccagatcagatcccaagaaacagaaacagaaacagaaacagaaacagaaacagaaacag 1680  
 Db 1681 AGCCAGATCAGATCCCAAGAAACGAAACGAAACGAAACGAAACGAAACGAAACGAAACGAA 1680  
 QY 1741 gatgagacag 1740  
 Db 1741 GATGACACAG 1740  
 QY 1801 ctggtgtcagtgacag 1800  
 Db 1801 CTGGTGTCAAGTGTGACAG 1800  
 QY 1861 gtcagatccatattatgag 1860  
 Db 1861 GTCAGATCCATATTATGAG 1860  
 QY 1921 caactgaatcagatattatgag 1920  
 Db 1921 CAACGTATATGACAGTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
 QY 1981 acccttaag 1980  
 Db 1981 ACCCTTAAG 1980  
 QY 2041 gcaag 2040  
 Db 2041 GCAAG 2040  
 QY 2101 ctgcccattgaag 2100  
 Db 2101 CTGCCCATTTGAAG 2100  
 QY 2161 gggattgagggcttgatgatgcgccaagagaatgctgcctcgaagataatgggt 2220  
 Db 2161 GGGATTGAGGGCTTGATGATGCGCCAAGAGAAATGCTGCCCTCAGTATGAGTAATGGGT 2220  
 QY 2221 gacacag 2280  
 Db 2221 GACACAG 2280  
 QY 2281 atagaag 2280  
 Db 2281 ATAGAAG 2280  
 QY 2341 tcccatagatgag 2340  
 Db 2341 TCCCATAGATGAG 2340  
 QY 2401 agtgaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2400  
 Db 2401 AGTGAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2400  
 QY 2461 gaaag 2460  
 Db 2461 GAAAG 2460  
 QY 2521 caattgagatcgttaaaacagaaacaaacaaacaaacaaacaaacaaacaaacaaacaaac 2520  
 Db 2521 CAATTGAGATCGTTAAACAGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2520  
 QY 2581 actaaactcaatcagtgatgaacagaaagagagagagagagagagagagagagagagagag 2580  
 Db 2581 ACTAAACTCAATCAGTGTGATGAACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580  
 QY 2641 tccctgagcag 2640  
 Db 2641 TCCCTGAGCAG 2640  
 QY 2701 actttgag 2700  
 Db 2701 ACTTTGAG 2700  
 QY 2761 tgcattgag 2760  
 Db 2761 TGCATTGAG 2760  
 QY 2821 gatgatgaag 2820  
 Db 2821 GATGATGAAG 2820  
 QY 2881 gactgtgatccacagcagatgacagcagcagcagcagcagcagcagcagcagcagcagc 2880  
 Db 2881 GACTGTGATCCACAGCAGATGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2880  
 QY 2941 aaccaaag 2940  
 Db 2941 AACCAAAG 2940  
 QY 3001 aagatag 3000  
 Db 3001 AAGATAG 3000  
 QY 3061 gggag 3060  
 Db 3061 GGGAG 3060  
 QY 3121 ataaataacag 3120  
 Db 3121 ATAAATAACAG 3120  
 QY 3181 gag 3180  
 Db 3181 GAG 3180  
 QY 3241 gctgaatcag 3240  
 Db 3241 GCTGAATCAG 3240





QY	181	cttgatgacattcttglytatgtgacgagacataaagacactggtgacagtgttat	240
Db	181	CTTGATGACATCTTTGTGATGTAGCAGACGATTAATAACAGACTGGTATGCAAGTGTGGAT	240
QY	241	gagcagatccacataccacggaagtgatagcaccagttgcagttccaggttacccagagc	300
Db	241	GAGCAGATCCACATCACGGAGGTGATGGACACAGTCCAGTTCCAGGGTACCAGAG	300
QY	301	ccagaatatttgtagtgagcttgcaccacaatgtctccagctcttaagcttacca	360
Db	301	CCAGAGTATTTGGTAGTGAAGCTTGGACCAACAAATGTCTCACCTTTACGCTTAACCA	360
QY	361	gcaacaagttaaattagatgcacacttcagttccctcgagcaaatatgtcctctcagt	420
Db	361	GCAACAAGTAAATTGATGAGTCACACCTTCAAGTCTTCGAGCAAAATATGTGCTTCATGTT	420
QY	421	cgacgcagtagtagaccacgtcctaattggcctctccactctgtgaagtatgaattt	480
Db	421	CGACGCAGTAGTAGCCACGACTTAATTGGCCTCTCCACTTCTGTCAAGTATGTAATTTT	480
QY	481	tctctgaagagccttcaagaaaaatcccaacgctgtgtcaacacacagttgctctc	540
Db	481	TCCTCGAAGAGCCTTCAAGAAAAATCCCAACGCTGTGTCAACACAGCTGTCTCTC	540
QY	541	aagcagaacactgtctgtagtctctaaacctgcgcagagaanaagatgaaactcaga	600
Db	541	AAGCAGAAACACTGTGTGGAGTCTCTAAACCTGCACAGAAAGATGAAATCAACGA	600
QY	601	agctcccccggagatactagtaactggtcctaacaattccagagaaactgtcgtcg	660
Db	601	AGCTCCCCCGGATACCTAGTAACTGGTCTCTAACCAATTTCCAGACAGCAATGTCTGCTG	660
QY	661	tctctgagtagccagtcaccacatggttggcgaagtgtgcggagaacaagaacagatagag	720
Db	661	TCTCTGAGTAGCCAGTCACCCCAATGGTGGGCAAGTGGCTGGAGAAACAAGACAGAGTAG	720
QY	721	gatggagaagaagagataaacaagtcggttgaaacctgttgacatgtgcacagtggttg	780
Db	721	GATGGAGACGAAGAGATMAACAGTCGTTGTAAACCTTTGGACATGCTGCACACGGTTTG	780
QY	781	gagcatalaccaaacttctcttgatgatatgttlaagctcgtlaagaagtcaccaacgat	840
Db	781	GAGCATATACCCAACTTTTCTGTGATGATATGTTAAAGCTCTGAAGAGTCCCAACGAT	840
QY	841	ggaagggcctcctgggaatccatctagtgctcttaagtgctcgaagcggacaccttggg	900
Db	841	GGAGGGCTCTGGGAATCCATGTAGTCCCTTTAGTGCTCGAAGGCGGACAAACCTGGGG	900
QY	901	ttattagtaaacgatttggaaagtgctgaagctgcgaacatgaanaatttttctgtgag	960
Db	901	TTATTAGTAAACGATTGGAAAGGTGTAAGCTGAACATGAANAATCTTTTCTGTGAG	960
QY	961	aatgatattgcttgaagataatgtagtgcgacctcgaaatagaagaattgaacaagra	1020
Db	961	AATGATTTGCAATTCACGAGTAAATGATGGCGACCTTCGAAATAGAAAGATTTGAACAAGA	1020
QY	1021	caacatactgttcgccaagccalgtctaacccaatcaattgttctcaagtgttccctga	1080
Db	1021	CAACATATGTTCGCCAACCCATGGGTACACCAATCATTTGGTTCCATGTGGTCTCTCA	1080
QY	1081	gcaaataaagcagatgtgaaactatccaaagttagaagaagaacattactattcaagc	1140
Db	1081	GCAAAATAAGACAGATGTGAACACTATCCCAAGTGAAGAAACAAATTACTATTCAAGC	1140
QY	1141	cgatttagccctgcgaagcagatatactgacaacaggaatgtgaacagtgcagggtctaac	1200
Db	1141	CGTTTAGCCCTGCAGACGAGTATTTTACAACAGAGATGTGAACAGTGCAGGGCTTCAC	1200
QY	1201	agagtcagagagacaccccgacttgaaacacccgcgttgacagatagactctcactcaaga	1260
Db	1201	AGGTCAGAGAGACACCCCGACTGAACACCCGCGCTTGAGCAGATAGACTCTCAAGA	1260
QY	1261	ctacctcatagcgacacccctcgggaaacacacatccgcgtccagctcggaacctcag	1320

[illegible]

Db 2332 TCCATGATGATGTTGGTTGTGACGCGCATGCTGTGACTTGGGCCAAGCTGCATTC 2391  
Oy 2401 agtattcagccagctgctcttgagtcacagatgttgaaccagcttcttgctttcaacga 2460  
Db 2392 AGTGATGAGCGGATGCTGCTTTGAGTCCAGATGTGATCCAGTTCCTTCTTTCAACGA 2451  
Oy 2461 gaagatttgacgttcacagatgttcacagaaaaacacacaaagcaatttcagatccagt 2520  
Db 2452 GAAGGATTTGGACGTCAAGATGTGCAAAAAACACAAACCAATTTTCAATCCCAT 2511  
Oy 2521 caattggatttcgttaaacacagaaatcaaaagcatggattttagatagctacag 2580  
Db 2512 CATTTGGATTTGCTTAAACAGAAATCAAAAACATGATTTAGGTTAGCTACAGAG 2571  
Oy 2581 actaaactcaatagatgtgatgacagagaagcaggttctccacagagatgttggtcct 2640  
Db 2572 ACTAAACCAATACAGTGGATGACAGAAAGAGTTCCTCCACAGAGATGTGGGTCT 2631  
Oy 2641 tccctggctcgaagaagtcagagctcgttgagagatctgcagaccgagcttgccaggtg 2700  
Db 2632 TCCCTGGGCTCGAAGACGTCAAGCTGTTGGAGAGTTCGACAGCCGCAAGTTCGAGGTG 2691  
Oy 2701 acttggaatggagatactcccttcacatcgtccagcgccggagataatgaagcagggga 2760  
Db 2692 ACTTGAATGGGATATTCCTTTCATGCTCCAGCGCCGGATATACAGAGCGAGGGA 2751  
Oy 2761 tgcagtgagcttcacagagctgcacacacaaatctatgataaaccccggtagatgat 2820  
Db 2752 TGCATGAGAGCTTCAGAGCTCCATCGCAATCTTATGATTAACCCCGGTACATGAT 2811  
Oy 2821 gatgatgaagcaltgagagactcttgagagaagacacagaagaagttcaaatcaaggaga 2880  
Db 2812 GATGATGAGAGGAGTGAAGACCTTGGAGAGACACAGAAAGAACTTCAAGATCAGGAGA 2871  
Oy 2881 gactgtgatccacagcagtgatcagccttcacactcctcgtgagagacaaatgatatga 2940  
Db 2872 GAGTGTGATTCACAGCCAGATATGAGCTTCCACTCTCTGAGAGACAAATGATGGA 2931  
Oy 2941 aaccaaagaaagtgtgataagaatcgtgataaaagaataaaactgtaaaagaagaag 3000  
Db 2932 AACCAAGAGAAAGTGATAGACTGATGAAAAAGATTAACCTGGAAGAAAGAAAGAG 2991  
Oy 3001 aaagtaagataaagagaagaataaaatgaagaagcagaagaagatgctgaaggtctg 3060  
Db 2992 AAAGTATAGAGTAAGAGAAAGATTAATGAAGCCCAAGAGGATGCTGAAGGCTTGG 3051  
Oy 3061 ggaagcagttcaggttctggaacacatcgaaaagaatgaagaatgagaacacgggttaa 3120  
Db 3052 GGAGACATGTTCAAGTTCGCTTGGCAAAATGCAAAAGATGACAAAGATTGAGAAAAGG 3111  
Oy 3121 aataaatacagaatccttcaatcagaagaagagagatacgaatgaagcagagcag 3180  
Db 3112 ATAAAAATACAGGAATCCCTTATCATCAGAGAGAGAGATGATGATGAAGCAGAGCGAG 3171  
Oy 3181 gagagatcacaagcacaacacgcgaagaatttaggagaagcagaagctcgagagcgtgactat 3240  
Db 3172 GAGAGATTTCAGACCAAAATCGAGAAATTTAGGAGAACGCAAGCTCGAGAGCTGACTAT 3231  
Oy 3241 gctgaatcagaatcttcatatcgagcatttggctgtgataaggttaattgtatgggga 3300  
Db 3232 GCTGAATTCAGATTTTCATCGACATTTGGCTGTGATGATGATTAATGATGAGGGA 3291  
Oy 3301 gttcttctatgaagttcagatggtctcagacgttagacactcagaagccacagaaggg 3360  
Db 3292 GTTCTTCTTATGAGAGTTCCATGCTGCTCAACGCTAGACCTCAAGCCCAAGAGAGG 3351  
Oy 3361 catatgatgatgctcttctatgagcccaagtcaagaagccgggaattccaacccctcaact 3420  
Db 3352 CATATGATGATGATGCTTTGATGCGCAAGTCAAGAGCCGCGGAATTCCAAAACCTCACCT 3411  
Oy 3421 gtagacagtaacagatacaactcctagcaatcagatcgatatacagcgtctgagcaaga 3480  
Db 3412 GTAGACATGATCAACATCTCTAGCAATCTGATCGGATACAGCGTCTGAGGCAAGAA 3471

Oy 3481 ttccagcaagcaaaagataagaatgtagaagatcgtcggcggacatagtttgaag 3540  
Db 3472 TTTTCAGCAAGCAAAACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3531  
Oy 3541 caaccttgccgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaag 3600  
Db 3532 CAACCTGGCCCAACGACGCGCGCGGACGACAGAGCGGCGACCTCGGTGTCCGTGGAG 3591  
Oy 3601 gtcagatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3660  
Db 3592 GTGCAATGACAGCGGACGCGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3651  
Oy 3661 tacagctctgctcctcggcaagcaggaatctgcaagctcgtctccagagctcttg 3720  
Db 3652 TACAGCTCTGCTCGCGCAAGCAAGAAATGCAAGCTCGGTCTCCAGAGCTCTTG 3711  
Oy 3721 gagcagaactactcccttgaggagagcttcagagatgccaagaagagacccagtaacc 3780  
Db 3712 GAGCAAACTACTCCCTCGGAGAGAGCTTCAGAGAGTCCAAAGAGAACCCAGGTACTCC 3771  
Oy 3781 agctacaaagctcagagagagcgtacactgagagacatggtttcaagcagaggtcatg 3840  
Db 3772 ACTTACAAAGCTCCAGAGAGAGCTTACCTGAGAGAGATGCTTCAAGCCAGGTCATG 3831  
Oy 3841 ctgaaactcagagagctccttcgccaagacagagcagagcagagcagagatgaagaag 3900  
Db 3832 CTGGAATCTCAGAGAGCTCTTCGCCAGAAACAGAGCGGAGAGAGAGAGATGAAGAAG 3891  
Oy 3901 cagctcctcctcggagggcccaagcactatgactgataagaagaatccagaagccag 3960  
Db 3892 CAGCTCTCTCGAGAGGCGCCAGCAACTATGACTCTATTAAGAAATCCAGAGACCCCACT 3951  
Oy 3961 taagcccccacaaagggcccttcggcaagatgtgccccccctccctcagagttgag 4020  
Db 3952 TACGCCCTCCCAAGGGGCTTCGGCAGAGATGTCCCCCTCCCTCTCAGGTTGG 4011  
Oy 4021 aggtgaacagagctcagactcctgagaagaaggagcccttctatctga 4071  
Db 4012 AGGCTGAACAGACTTCAGACTCTTGAGAAAGAGAGGAGGCTTCATTCCTGGA 4062  
  
RESULT 4  
AF467004 3960 bp mRNA linear PRI 07-FEB-2002  
LOCUS  
DEFINITION Homo sapiens partitioning-defective 3 protein splice variant d  
(PARD3) mRNA, complete cds, alternatively spliced.  
ACCESSION AF467004  
VERSION AF467004.1 GI:18568347  
KEYWORDS  
SOURCE  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Gao,L., Macara,I.G. and Joberty,G.  
TITLE Multiple splice variants of PARD3 and of a novel related gene,  
PARDL, produce functionally different proteins  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3960)  
AUTHORS Gao,L., Macara,I.G. and Joberty,G.  
TITLE Direct Submission  
JOURNAL Submitted (10-JAN-2002) Center for Cell Signaling, University of  
Virginia, PO Box 800577, Charlottesville, VA 22908, USA  
FEATURES  
source  
1..3960  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..3960  
/gene="PARD3"  
/note="PARD3"  
1..3960  
/gene="PARD3"



QY 1501 cagatgcccgaacttaagcagagagacagacttatagaglaaatgagtagatgag 1560  
DB 1501 CAGATGCCCGCACTTAAAGCAGAGACAGACTTATAGAGTAAATGAGATGATTAGG 1560  
QY 1561 ggcgaatcccaagaggaagttgcttgcgtgtgttgagaagacacccaagaatgagaagactgtg 1620  
DB 1561 GGCgaatcccaagaggaagttgcttgcgtgtgttgagaagacacccaagaatgagaagactgtg 1620  
QY 1621 agccttcggtccttcgcagagaagccttcacccaaggaactgaatgacagagca 1680  
DB 1621 AGCCTTCGgtccttcgcagagaagccttcacccaaggaactgaatgacagagca 1680  
QY 1661 agccagatgagatcccaaaagaaacgaagcagaagaatgagatatgtcttcacact 1740  
DB 1661 AGCCAGATGAGATTCcaaaagaaacgaagcagaagaatgagatatgtcttcacact 1740  
QY 1741 gatgacacaaaggaattctcgaacttgaagtcccaactaatgatacagatcagagc 1800  
DB 1741 GATGACACaaaggaattctcgaacttgaagtcccaactaatgatacagatcagagc 1800  
QY 1801 ctgtgtcagtgatcaaaagatcaaccggtcaaaagaaacacacagactttgggaactctt 1860  
DB 1801 CTGTGTCAgttgatcaaaagatcaaccggtcaaaagaaacacacagactttgggaactctt 1860  
QY 1861 gtcaagttcatatgaatgagagagacacatctaaagaatgaaagccttcgagtgagat 1920  
DB 1861 GTCAAGTTCAatgaatgagagagacacatctaaagaatgaaagccttcgagtgagat 1920  
QY 1921 caactgataagtaaatgaggaatccctgtttggcagaagacaaaccaaagatgacatgaa 1980  
DB 1921 CAactGATAagtaaatgaggaatccctgtttggcagaagacaaaccaaagatgacatgaa 1980  
QY 1981 accctaagaaggtctatgtctactgaaggcaataaagaggaatgatacagcttatgtc 2040  
DB 1981 ACCCTAAGAAGGTCTATgtctACTgaAGCAATAAAGAGAAATATCCAGCTTATTGT 2040  
QY 2041 gcaaggaataaagcaagtgcaatgagtgaaatcaactggagaccccttgagactgag 2100  
DB 2041 GCAAGGAATAAAGCAAGTGCATGCAATGAGCTGAAGTCACTGGAGGCCCTTGAGACTGAG 2100  
QY 2101 ctgcccattgaagaacgcttgatgatatagagaagaagaattcccatctccctcaagct 2160  
DB 2101 CTGCCCATTGAaAGCCTTGATGATAGAGAAACGAAGAAATTTCCCATTCCTCAACAT 2160  
QY 2161 ggaattgagggcttgatgaaatcgccagcagaagaatgctgcccctagtagaataagtg 2220  
DB 2161 GGAATTGAGGGCTTGATGAAATCCCCCAGAGAAATGCTGCCCTTAGTAGAGTAATGGGT 2220  
QY 2221 gagtcaaglaaataacacagctgtccctcaagtgaaatgaccccaagatgacactgcat 2280  
DB 2221 GAGTCAAGTAAATACCACTGTGCCCTACAGTGAATATGCCCAAGATGACACTGCAT 2280  
QY 2281 atagaagatgacaggttccaggtgtcttccacacatctctcgacacagctccctccagc 2340  
DB 2281 ATAGAAGATGACAGGTTCAGAGTGTCTCCACATCTCTGACACAGCTCCCTTCCACG 2340  
QY 2341 tcccatgatatgtggtttgtgacgagcagaatgctgtaacttgggccaagctgcaatc 2400  
DB 2341 TCCCATGATGATGTGGGTTGTGACGGCAGATGCTGTGACTTGGGCCAAGCTGCAATC 2400  
QY 2401 agtgaatcaagcagactgctcttgatccagaatgltgacagactctgtcttcaaga 2460  
DB 2401 AGTGAATCAAGCCGACTGCTTGTGAGTCCAGATGTGATCCAGTCTTTCACAGCA 2460  
QY 2461 gaaagattgagacgtcaagatgctcagaanaacgacaagaactttcaaatgccagt 2520  
DB 2461 GAAAGATTGAGAGCTCAGATGTGTCAGAAAAACGACAAACCAATTTTCAATGTCAGAT 2520  
QY 2521 caattggaattcgttaaaacacgaataatcaaaaacatggaatttagatagctgacag 2580  
DB 2521 CAATTGGAATTTCGTTAAAAACGAAAAATCAAAAAGCATGGAATTTAGTATAGCTGACGAG 2580

QY 2581 actaaactcaatagatgagatgacagaagaacaggttctccacagagatgtgggtcct 2640  
DB 2581 ACTAAACTCAaATACAGTGGATGACCAGAAAGCAGGTTCTCCACAGAGATGTGGGTCT 2640  
QY 2641 tccctggtctgaagaagtcaagctcgtttggaatctcgcagaccgagcttgccaggtg 2700  
DB 2641 TCCCTGGTCTGAaAGCAAGCTCGTTGGAGACTCTGCAACGCCAGTTGCCAGAGTG 2700  
QY 2701 acttgaatgggaatcttccatctgcacgtccagcccgagtaatacgaagcagggga 2760  
DB 2701 ACTTGAATGGGAGATTTCTTCCATCTGTCACCGCCCGGATATATCAGAGCCAGGGGA 2760  
QY 2761 tgcgaatgagactcaagatgacatcgacaaacttatgataaaccggttagatgat 2820  
DB 2761 TGCgaATGAGACTTCAGAGCTCCATCGCAATCTTTATGATTAACCCCGGTAGATGAT 2820  
QY 2821 gatgtgaagatgagagaccttggagaagaacacagaagaagaatgatacaggggga 2880  
DB 2821 GATGATGAAGGATGAGACTTGGAGAGAACACAGAAAGAAATTCAGATCAGGGAGA 2880  
QY 2881 gagtctgtacacagcagtgatcagccttcacactctgtgagagacaaatgatatga 2940  
DB 2881 GAGTCTGTATCCACAGCCAGCTGATCAGCTTCCCACTCTCTGAGAGAACAAATGATGA 2940  
QY 2941 aaccagaagaagtgatgaagaactgataagaataaagaatlaaaactggaagaagaagaag 3000  
DB 2941 AACCAAGAGAAAGTGATGAGACTGTAGAAAAAAGATTAACCTGCAAAAAAGAAAG 3000  
QY 3001 aagaatagagataaagaagaagaataaagaagaagaagaagaagaagaagaagcttg 3060  
DB 3001 AAGATAGAGATAGAGAGAGATTAATGAAGCCAAAGAGGAATGCTGAAGGGCTTG 3060  
QY 3061 ggaagcatgttcaagtttggcaaacatcgaaaagaatgaaagaatgagaagaacggtlaa 3120  
DB 3061 GGAGCATGTTC----- 3072  
QY 3121 ataaaaatacaggaatccttcatcagaagaagaagaatagaaatgaaagcagagcag 3180  
DB 3073 ----- 3072  
QY 3181 gagaagattcaagccaaactcgaagatttaaggagaacgaagctcgaagctgactat 3240  
DB 3073 ---AGATTCAAGCCAAACTCGAGATTTAGGGAACGAAGCTCGAAGGGTGCAT 3129  
QY 3241 gctgaatcaagatttcatcgcgaacttggctgctgagatgagatgaatgatatgggga 3300  
DB 3130 GCTGAaATCAAGATTTTCAATCGACATTTGGCTGTGATGATGAGTAATGATGGGGA 3189  
QY 3301 gttcctcttataagagttcatalgctcacaagctagaacctcagaagcccaagagaag 3360  
DB 3190 GTTCTCTTATGAAGTTCCATGGCTCTCAACGCTAGACCTACAGGCCAGAGAAAGG 3249  
QY 3361 catatgattgagcttttgatctgccaagtcagaagaacgaggaattccaaacctact 3420  
DB 3250 CATATGATGAGCTTTGTATGCCCCAAGTCAAGAAACCCCGGAATTCAAACCTCTACT 3309  
QY 3421 gtagaagtaacagatcaactcctagcaatcatalgcatcagatacaggtctgagcaaga 3480  
DB 3310 GTAGACAGTAACAGATCACTCTAGCAATCATGATCGGATACAGGCTGTGAGCGAAGA 3369  
QY 3481 ttccagaagaacgaagaatgaagaatgataagatcgtcggcgagactatgtttag 3540  
DB 3370 TTTTCAGCAAGCAAAAGCAAGATGAAGATGTGAAGATCGTCGGCGGACCTATAGTTTGAG 3429  
QY 3541 caaaccttgccgaacgacgagccgagagcagaagcgcagacactcgtgtccgtggag 3600  
DB 3430 CAACCTTGCCGAACGACAGCGCCGGGACGACAGCGGGCGAATCTGGTGTCCGTGGAG 3489  
QY 3601 gtgcagatgcagcgcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3660  
DB 3490 GTGCAAGATCAGCGGAGCGGAGCGAGAGCGAGAGAGCTCCACAGAGGCCAGCGCCAG 3549  
QY 3661 tacagctctcgtcgcgcaagaacgaagaataatgccaagctcggtctccaggaactcttg 3720

misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature

BASE COUNT      1164 a            899 c          1049 g            821 t

ORIGIN

misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature

BASE COUNT      1164 a            899 c          1049 g            821 t

ORIGIN

```
QY 601 agctcccgcgagatactagtaactggtcctaaaccaattcagagagacaatgctgcgtcg 660
Db 601 AGCCCTCCGCGGAGTACTAGTAACGTGCTTAACCAATTTTCAGAGAGACAAATCTCGCTCG
QY 661 tctctgagtgacagtcacccaatggtgagcaagtgctgagaaacaagaatgag 720
Db 661 TCTCTGAGTGCGAGTCACCAATGAGTGCGCAAGTGCGTGAACCAACAGAGTGAAG
QY 721 gatgagacagaagagatacagtcgtgtgaaacctgttgacatgctgacacgggttg 780
Db 721 GATGGAGACAGAGAGATATACAGTGTGTGAACTGTTGAGACATGCTACACAGGGTTTG
QY 781 gagcatataccaactcttctctgagatgatatgtaaaagctgcgtagaagtcaccaagat 840
Db 781 GAGCATATACCAACTCTTCTGATGATATGTAAGCTGTGAAGTCTGCCCAACGAT
QY 841 ggaagagctctgagatcaatgtagtgccttcaagtgcgcgagggagagaaacctggag 900
Db 841 GGAGGGCTCTTGAGATTCATGTAGTGCCTTCAAGTGCAGAGCGGCAAGACCTGGGG
QY 901 ttattagtaaaaagcttgagaaaggtgtaaaagctgaaatgaaatcttctcgtag 960
Db 901 TTATTAGTAAGAAAGATGAGAAAGGTGTAAGCTGAACATGAATCTTTTGTGTAG
QY 961 aatgattgcatgtcagaagtaataatgagcgacctcgaaatagaagattgaaacaaga 1020
Db 961 AATGATTGCATTTGTAGAGTAATATGATGCGCACTTCGAATATGAAATGAAATTAACAAACA
QY 1021 caacatattgttcgcaaaagcctgtacacacatcattgtgttcattgttcctgta 1080
Db 1021 CAACTATGTTTTCGCCAAGCCATGCGTACACCCATCATTTGTTTCATTTGTTCTGTGA
QY 1081 gcaaaataaagacagtaataaactatcccaagtgagaaagaaactatctatctaacg 1140
Db 1081 GCAAAATAAGACAGATATGAAACAACTATCCCAAGTGAAGAAACAAATTAATTCAAAC
QY 1141 cgtttagcctcgacagcagcagatataatgacaacagagatgtgtaaacgtagcaggctcac 1200
Db 1141 CGTTTACCCCTGACAGCGCATATATGTAACAGAGAGTGTAAACAGTCCACGCGCTTCAAC
QY 1201 acggtgcagagagaccccgactgaaacacgcgctgagcagatagactctaacataaa 1260
Db 1201 ACGGTGCAGAGACACCCCGACTGAACACCGCTGACAGATTAATCTCATCTCACTAACA
QY 1261 ctacctctatgcgcacacccctcgagaaacacacatccgctcgcacgctcgacactcag 1320
Db 1261 CTACCTCATAGCGCACACCCCTCGGAAACCAACATCCGCTCCAGCCCTCGCACCTCAG
QY 1321 aatgattatgtagacagcgttaagcagtggtataaacaacaaaataagcagaagctt 1380
Db 1321 AATGATTATGATACACTGTAAAGCAGTGTATTAACACCAAAAATAATAGCAAGAGCTT
QY 1381 aatacagagcttaagaagagtaacagaagtttggattcagatacactccagaatgta 1440
Db 1381 AATACTCAGCTTAAGAAAGTACAGAGTGTGGGATTCAGATCACTTCCAGAGATGTA
QY 1441 acaataagtggtcagctccaaactctatgtaaaaaaatctcccccggggggcgccatt 1500
Db 1441 ACAATAAGTGGCTCAGCTCAATCTATGTAAGAAACATTTCCCGGGGGGGCGGCATT
QY 1501 cagatagcagcagcttaagcagagagacagacttataagagtaaatgagtagattagtg 1560
Db 1501 CAGATAGCGCGCACTTAAGCAGAGACAGACTTAATAGAGTAAATGAGATTAATTAGTG
QY 1561 ggaacaatcccaagagagagtgcttcgcgtgtgagaaagacaaagatgaaagaaactgtg 1620
Db 1561 GGCAATATCCCAAGAGAGAGTGTTCGCTGTGAGAGACCAAGATGGAAGAACTGTGTG
QY 1621 agcctcttggtcttcgcagagaaagacgcttccaccaaggagaaatgtagagca 1680
Db 1621 AGCCTTCTGCTGCTTTCGCGCAGAGAGACGCTTCCACCCCAAGGAGAACT
QY 1681 agcagatgacagattccaaaagaaagcagaagatgagagataatgttcttaacact 1740
Db 1681 -----GAAAGCAGAAAGATGAGATATGTTTAAACCTT 1701
QY 1741 gatggaccaggaatcttgcacatttgaagttcccaacttaataatcagaatgtaggc 1800
Db 1702 GATGGACCAAGGAAATTTTGACATTTGAAGTCCCACTTAATATTAAGATGTGAGGC
QY 1801 ctgtgtcagtgctcaaaaglaacccggtcacaagaagaaacagcagatttggaaatctt 1860
Db 1762 GTTGTGTCTAGTGTCAAGGTAAACGGGTCAAAAGAAACACGACCAATTTGGAACTT
QY 1861 gtcaagtcacattataatggaagagcagacatctaaagatggaagcttcgggtgaaatgat 1920
Db 1822 GTCAAGTCCATTAATATATGAGAGAGACATCAAGATGGAAGCTTGGGTGAATGAT
QY 1921 caactatagcagtaaatgagaatccctgttggagaaacaaacaaagatgtaggaa 1980
Db 1882 CAACATATAGCACTTAATATGAGAGATCCCTGTGGGCAAGCAACAAAGATGCAATGAA
QY 1981 accctlaagaaggtctatgtctactagaagcaataaagcagaatgataccagcttatgtt 2040
Db 1942 ACCCTAAGAAGGTCTATGTCTACTGAAGGCAATAAACGAGATGATCCAGTATTTGTT
QY 2041 gcaaggaagaataagcaagtgcacatgagctgaagtcacctggagacccctggacctgag 2100
Db 2002 GCAAGGAGATTAAGCAAGTGCAATGAGCTGAACCTGAGGAGCCCTCGACCTGAG
QY 2101 ctgcccattgaaacagcgttggatgtagagaagcagaagaattcccatccctcagat 2160
Db 2062 CTGCCATTGMAACACCGTTGGATGTATGAGAACGAAATTTCCATTCCCTTCAAGT
QY 2161 gggattgagggccttgatgaatcgcacagcaagaatgctgcctcagtagaataatgggt 2220
Db 2122 GGGATTGAGGGCGCTTATGATATCCCCAGAGAAATGCTGCCCTCAGTAGAGTAATG--- 2178
QY 2221 gagtcaggttaataaccagctgtcccttaacagtgtaataatgcccagaatgaactgtcatt 2280
Db 2179 -----GGTAAATACCAAGCTGTCCCTACAGTAAATATGCCCAATGACACTGTCAAT
QY 2281 atagaagaatgacaggtgtgcagtgcttccctcacaactctctgaacagctctcttcacg 2340
Db 2233 ATAGGAAGATGACAGGTGGCAGTGTCTCTCCACATCTCTTCAACAGTCTTTCACG
QY 2341 tcccatgatatgtgaggttltgtaacgcagatgctgtaacttgggccaaggtctgaatc 2400
Db 2293 TCCCATGATGATGTGGGGTTTGTGAGCGGAGAGTGTGTACTTGGGCCAAAGCTTGCAATC
QY 2401 agtgaatcagcgcagctgcctttagatccaagatgtgataccagttctgtcttcaagaa 2460
Db 2353 AGTGAATTCAGCGACATGCTCTTGAAGTCCAGATGTGATCCAGTTTGTCTTCAACGA
QY 2461 gaaagattcgagctcagagatgtcagaaaacgcacaaagaatttcaagtgcagtt 2520
Db 2413 GAAGGATTTGGAGCTGAG----- 2430
QY 2521 caattggaattcgtlaaaacacagaaatcaaaaagacatgattagtagtcagag 2580
Db 2431 -----ATAGCTGACGAG 2442
QY 2581 actaaactcaatacagtgatgatacgaagaagcaggttctcccaagcagaatgtggctct 2640
Db 2443 ACTAAACTCAATACAGTGTATGACAGAAAGCAGGTTCTCCAGCAGAGATGTGGGTCTCT
QY 2641 tccctgtgctgaagaagtcacagctgttggagaagcttcagacgcgacgttgcgcaggtg 2700
Db 2503 TCCCTGTGCTGAAGAGATCAAGCTGTGTGAAGTGTGAGAGTCCCAAGTGTGCGAGGTG
QY 2701 acttgaatgaggataatctcttccatcgtccacggcgcgagataatccagaaagcagaagga 2760
Db 2563 ACTTTGAATGGGATATTTCTCTTCCATGTCACGCGCGGATATATCAGAGGCAAGGGA
QY 2761 tgcataatgagcttcagagctgcacatgcacaaacttataatgaacccgcggtatagat 2820
Db 2761 ----- 2820
```



```

Db 2623 TGCATAGAGCTTCAGAGCTGCATCGCAATCTTATGATAAACCCGGGTAGATGAT 2682
QY 2621 gatgatgaagcattgagagccttggagaagaacacagaagaagttaagaatcagaagaga 2880
Db 2683 GATGATGAAGGATGAGAGCTTGGAAACAGACACAGAAAGCTTCAAGATCAGGAGA 2742
QY 2681 gactgtatccacagcagatgatalcaagccttccacactcttggagaagacaaatgata 2940
Db 2743 GAGTGTATCCACAGCCAGATGATCAGCTTCCACTCTCTGTGAGACAAATGAAATGAA 2802
QY 2941 aacaaagaagaaggatgaataagcgtatgataaagaagaataaactggaagaagaaga 3000
Db 2803 AACCAAGAGAAAGGTGATTAAGCTGATAGAAAAAGATGAAAACTGAAAAAGAAAGAG 2862
QY 3001 aaagatagaagaagaagaagaataaataagaagcaagaagaagaatcttgaaggcttg 3060
Db 2863 AAAAGATAGAGTAAAGAGAGATGAAAAAGAAAGCAAGAAAGCAAGTCTGAAGGCTTG 2922
QY 3061 ggaagacttcaagcttggcaaacatcgaagaatgaagaatgaagaagaaggatgaaga 3120
Db 2923 GAGACATGTTGAGCTTGTGGCAAAACATGCAAAAGATGCAAGATGAGAAAAAGGTTAA 2982
QY 3121 ataaataacagaagatcccttaccatcagaagaagaagaagaataagaagaagaagcag 3180
Db 2983 ATTAATAATACAGAGATCTTTACATCAAGAAAGAGAGATACGATGAAAGCAGAGAG 3042
QY 3181 gagagattcaagccaaactcgaagaattagaagaagaagaagaagaagaagaagaaga 3240
Db 3043 GAGAGATTCAGCAACCAACCTCAGAAATTTAGGAAACGAAAGCTGAGACCGTGCAT 3102
QY 3241 gctaaattcaagaatttccacagacatttgcctgctgctgctgctgctgctgctgct 3300
Db 3103 GCTAAATTCAGAAATTTTCATCGACATTTGCTGTGATGATGATGATGATGATGATGAT 3162
QY 3301 gcttctctatgaagtttccacagcctcgaagcctcgaagcctcgaagcctcgaagc 3360
Db 3163 GTTTCTCTTATGAGAGTTCCATGCTGTCAACCTGAGACCTGAGACCTGAGACCTGAG 3222
QY 3361 catatgatgatgcttctgtatgcccagaatcagaagaagaagaagaagaagaagaaga 3420
Db 3223 CATATGATGATGCTTGTGTATGCCCCAAGTCAAGACCGCGGAAATTCAGAAACCTC 3282
QY 3421 gtagagcattcaagaatcactcactcactcactcactcactcactcactcactcactc 3282
Db 3283 GTAGACATGATCAAGATCAATCACTCAGCAATCAGATGATGATGATGATGATGATGAT 3480
QY 3481 ttcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 3342
Db 3343 TTTTACAGCAAGCAAGCAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 3540
QY 3541 caaccctggcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3402
Db 3403 CAACCCGCGCCGAGACGACGCGCGCGAGCAGCAGCGCGCGACACTGCTGCTGGAG 3462
QY 3601 gtccaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3660
Db 3463 GTTCACATGATCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCG 3522
QY 3661 taacagctctgctcctcagcagcagcagcagcagcagcagcagcagcagcagcag 3720
Db 3523 TACACCTCTGCTCGCGCAAGAGCAAGAAATGCAAGCTGCTGCTGCTGCTGCTGCTG 3582
QY 3721 gagcagcactctcctcctggggaagcttccagatgagcagcagcagcagcagcagc 3780
Db 3583 GAGGAGAACTACTCCCTCGGGAAGCTTCCAGACTGCCAAAGAAACCCAGGACTATCC 3642
QY 3781 agctacaagaagctcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3840
Db 3643 AGCTACCAAGAGCTCAGGAACGCTTACCTGGAGAGACATGCTTCAACGCCAGAGCTC 3702
QY 3841 ctggaactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3900
Db 3703 CTGGAAGCTCAGAGAGCTCTTCCAGGAACAGAGCGGGAAGAGAGAGAGAGAGAGAG 3762

```

```

QY 3901 cagctccttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3960
Db 3763 CAGCCTCTCTCCAGAGGCGCCAGCAACTATGACTCTGATTAAGAAAGTCCAGAGCCAGT 3822
QY 3961 taagcctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 4020
Db 3823 TAGGCGCTCTCCAGAGGCGCCCTCCGCAAGATGAGTCCCTCCCTCTCTCAGAGTGG 3882
QY 4021 agcttgaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 4071
Db 3883 AGGCTGAACAGACTTCAGACTCTCTGAGAAAGAGAGGCGCTTCTATCTCTGA 3933

```

```

RESULT 6
AF332593
LOCUS 3822 bp mRNA linear PRI 16-NOV-2001
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein long
ACCESSION AF332593
VERSION AF332593
KEYWORDS variant b mRNA, complete cds, alternatively spliced.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3822)
AUTHORS Fang, C.M. and Xu, Y.H.
TITLE Down-regulated expression of atypical PKC-binding domain deleted
JOURNAL Cell Res. 11 (3), 223-229 (2001)
MEDLINE 21520080
PUBMED 11642408
REFERENCE 2 (bases 1 to 3822)
AUTHORS Fang, C. and Xu, Y.
TITLE Exon/Intron Structure And Splicing Variants of A Novel Human
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 3822)
AUTHORS Fang, C. and Xu, Y.
TITLE Direct Submission
JOURNAL Submitted (29-DEC-2000) Laboratory of Molecular and Cellular
OncoLOGY, Institute of Biochemistry and Cell Biology, Shanghai
Institutes for Biological Sciences, Chinese Academy of Sciences,
320 Yue-Yang Road, Shanghai 200031, P. R. China
FEATURES
source
1..3822
Location/Qualifiers

```

```

FEATURES
source
1..3822
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
1..3822
/note="ASIPLB: alternatively spliced; similar to Genbank
Accession Number AF196185"
/codon_start=1
/product="atypical PKC isotype-specific interacting
protein long variant b"
/protein_id="AAK69193.1"
/db_xref="GI:14579311"
/translation="MKTVVCGRTVRYVPCGDGIMKVFSLIDQAVTRYKATAPNT
WIOVHRLEHGDDGILDDIDLDVADKDLVAFEDDPHGGGDTASSGTOSE
ITGSELGNNVSAFOPOATSEIVTSVLRAMPVLRSSDPALICSTSVSDNSP
NSLSASHPNPKRWSTTGAETAKONTAGSPKCDKRDENRSLPRTSMKMOFORDA
VNDGPGIIVVFPFSKGRITGLVKRLKLEGGCAEHNTFREDDICVRIINDGLN
SVNSAGHTVORARLNHPPEQIDSHRSLPASHAHSOSEKNNTYSRSRPSQYIDNR
DRLIEVGVLDVGRSOEYVGLRSTKMGCTVSLVFEQEDAPFRLAKAEDDLYIVP
PDGAREFLTEVPLNDGSGAGIVYKRSKSEHNADGIVKSIINGAASQDGLR
VNDLIVAVNESILGKTNDAMETVRSKSTDESKRGMIOIIVARISKNKSKSGS
PPGPELPIETALDEGRKSHSITGISTGIESSRNAALSRIMGYOISLPTNMKDP
DTVIEDRLPVLPHLSQSSSSHDVDGVTADACTWAKAALISDACSLSPPDVP

```

```

CDS
1..3822
/note="ASIPLB: alternatively spliced; similar to Genbank
Accession Number AF196185"
/codon_start=1
/product="atypical PKC isotype-specific interacting
protein long variant b"
/protein_id="AAK69193.1"
/db_xref="GI:14579311"
/translation="MKTVVCGRTVRYVPCGDGIMKVFSLIDQAVTRYKATAPNT
WIOVHRLEHGDDGILDDIDLDVADKDLVAFEDDPHGGGDTASSGTOSE
ITGSELGNNVSAFOPOATSEIVTSVLRAMPVLRSSDPALICSTSVSDNSP
NSLSASHPNPKRWSTTGAETAKONTAGSPKCDKRDENRSLPRTSMKMOFORDA
VNDGPGIIVVFPFSKGRITGLVKRLKLEGGCAEHNTFREDDICVRIINDGLN
SVNSAGHTVORARLNHPPEQIDSHRSLPASHAHSOSEKNNTYSRSRPSQYIDNR
DRLIEVGVLDVGRSOEYVGLRSTKMGCTVSLVFEQEDAPFRLAKAEDDLYIVP
PDGAREFLTEVPLNDGSGAGIVYKRSKSEHNADGIVKSIINGAASQDGLR
VNDLIVAVNESILGKTNDAMETVRSKSTDESKRGMIOIIVARISKNKSKSGS
PPGPELPIETALDEGRKSHSITGISTGIESSRNAALSRIMGYOISLPTNMKDP
DTVIEDRLPVLPHLSQSSSSHDVDGVTADACTWAKAALISDACSLSPPDVP

```



VLARORRGSGFROINDERTKLMNVDOORAKGSPBROPSGLIKTKSSLTSLIOTPAVAEYTL  
 NGDI.PORHRRPRTIRIRGCGNBSFMAALIDKSTDKRAVYDDDEGNTLEEDTPESSRSR  
 ESVSTASDQPSHSLERONGNOEKGDITDRKDDTGKKEKROREKEDMMKAKGMLTK  
 GLGWMFIQAKTREFREFOAREYALIOHFRFGCDDELMYGSVSYESPMALNLR  
 POSREGHIMDALTAQVKKPAPNSKPSPVDSNRSTPSNHDRIQRLROEFOAKODEVRE  
 DRRATYSFEQWPNARPAPRSGSRHSVEVYOMQOROEERSSGOAORVSLSPERORS  
 KNASSVSODSWEQNYSPGECGOSAKEMPRYSYGSNGNYLGGGAFARVALTJEOJLOR  
 ROEORRKEQOQKQPPSEBGSNNYDXYKVDOPPSYAPKGPGRDOPVPPSSQVATLNLID  
 QTPERKGRFYS"

BASE COUNT 1114 a 885 c 1019 g 804 t  
 ORIGIN

Query Match	62.08;	Score 3529.8;	DB 9;	Length 3822;
Best Local Similarity	93.88;	Pred. No. 0;		
Matches 3820; Conservative	0;	Mismatches 2;	Indels 249;	Gaps 4;

OY	1	atbaagagacgctgacttcggaagaaacccgggctgctgcgctgaggagagccac	60
Db	1	atgaatgaagacccgctgctgctggaagaaacccgggctgctgcgctgaggagagccac	60
OY	61	atgaagcttctcaagcctcatccagcagcgctgacccgctacccaggaagccatcccaag	120
Db	61	atgaatgcttccagcctcatccagcagccgctgacccgctacccaggaagccatcccaag	120
OY	121	gatacaaacctactgtgatacaagctgcatcgcttggaaatctggagatggaatactagac	180
Db	121	gattccaaactactgattgatacaagctgcatcgcttggaaatctggagatggaatactagac	180
OY	181	ctttgtagcatctctttgtgattgtagcaagataaagaacagactgttagcagtgttat	240
Db	181	ctttgtagcatctctttgtgattgtagcaagataaagaacagactgttagcagtgtttat	240
OY	241	gagcagagatcccaatcaatcagaaggtatgacacaaigtgcacgttccacaggtatacccaagc	300
Db	241	gagcagagatcccaatcaatcagaaggtatgacacaaigtgcacgttccacaggtatacccaagc	300
OY	301	ccagagataattgtgtagtgtgagcttggcaaccaaatgtctcagccttcaagccttaacca	360
Db	301	ccagagataattgtgtagtgtgagcttggcaaccaaatgtctcagccttcaagccttaacca	360
OY	361	gcaaacagtgaatgtgaggtgacacccctcagtccctctggaagaaataigtcttcaatt	420
Db	361	gcaaacagtgaatgtgaggtgacacccctcagtccctctggaagaaataigtcttcaatt	420
OY	421	cgaacgcaatgtagcccaagcctctaatitggcctctccactctcgtlcaagatagtaattt	480
Db	421	cgaccacagatgtagcccaagcctctaatitggcctctccactctcgtlcaagatagtaattt	480
OY	481	tcctctgaaagagccttcaaggaataatcccaacagctgtgtcaacaaagcttgcctctc	540
Db	481	tcctctgaaagagccttcaaggaataatcccaacagctgtgtcaacaaagcttgcctctc	540
OY	541	aagcagaacacatgctgctggagctctctaaaacctcgacaggaagaagaatgaaactacaga	600
Db	541	aagcagaacacacatgctgctggagctctctaaaacctcgacaggaagaagaatgaaactacaga	600
OY	601	agccctcccgcggaatactagtaactgtgtcttaacccaatttcagagagacaatgtctgcctg	660
Db	601	agccctcccgcggaatactagtaactgtgtgtcttaacccaatttcagagagacaatgtctgcctg	660
OY	661	tctctgagtgccaagtcacccaatgtgtggcaagtgtgcgggaataaagaacagaatgag	720
Db	661	tctctgagtgccaagtcacccaatgtgtggcaagtgtgcgggaataaagaacagaatgag	720
OY	721	gattggaacaaagagataacagctgcttgaaccctgttggacatgctgcacacaggtttg	780
Db	721	gattggaacaaagagataacagctgcttgaaccctgttggacatgctgcacacaggtttg	780
OY	781	gagcatatacccaactcttctctgagatatalgyltaaagctcgttagaagttcccaagat	840
Db	781	gagcatatacccaactcttctctgagatatalgyltaaagctcgttagaagttcccaagat	840

QY	841	ggagggcctctgggaatccatctgtatgtgccttcaatgtctcgagggcgagaaacctggg	900
Db	841	GGAGGGCCTCTGGGAATCCATGTATGTGCTTCCTTCAGTGCTCGAGGGCGGAGAACCTCGGGG	900
QY	901	ttattagaaaaagatctggggaagagtggtgaagcgtgaacatagaatctttctgtag	960
Db	901	TTATTAGTAAAAAGATCTGGGGAAGAGTGGGATGAAGCTGAACATGAAATCTTTTCTGTAG	960
QY	961	aatgattcattctgcaggaattaaatgatgatgagaccttcgaatatagaatttgacaaga	1020
Db	961	AATGATTCATTCGTCAGGATTAATGATGGGACCTTCGAATATGAAGATTGAACAGCA	1020
QY	1021	caaatatgttctgcgaagccatgctgaacccatcaatttgttccatgtgtctctgca	1080
Db	1021	CAACATATGTTTCCGCAAGCCATGCTGACCCATCATTTGGTTCCATGTGTTCTCTGCA	1080
QY	1081	gaaataaaggagcgctltgaaacactatcccaagtggaagaacaattactatcaagc	1140
Db	1081	GAAATAAAGGACGTAATGAACAATCTTCCCAAAGTGAAGACCAATTCATTCAAGC	1140
QY	1141	cgctttagcccttgcaagccagatatttgacaacaggagtgtgaacqtlgcaaggtctac	1200
Db	1141	CGTTTACCCCTGACAGCAGTATATGTGACAAAGAGTGTAAACAGTGCAGGCTTCAAC	1200
QY	1201	acgyltgcagagagacccccgcatctgaaccaacccgcttgagcagatagactctcaataa	1260
Db	1201	ACGGTGCAGAGAGACACCCCGACTGAACCAACCCGCTGAGAGATGATCTCATCAAGA	1260
QY	1261	ctacctatagcgcacacccctcgggaaaaacccatcgcgtccagctctggcaacctag	1320
Db	1261	CTACCTCATAGCGACACACCCCTCGGAAAAACAACATCCGCTCCACCCTCGGCACCTCAG	1320
QY	1321	aatgtatttagtaagcttgtaagagtggtlataacacaaaaaataagcgaagagctt	1380
Db	1321	AATGTATTTAGTACGACTGTAAGCAGTGGTTATATACCAAAAAAATAGCAGAGGCTT	1380
QY	1381	aatatccagcttaagaaagttacagaaggtlttggaatcagcatcaacttccagagatga	1440
Db	1381	AATATCCAGCTTAAAGAAAGGTACGAAGGTTGGATTCAGATCATCTCCAGAGATGTA	1440
QY	1441	acaataagttgctcaagctccaactatgtgaaaacaacttcctcccggggggcgccat	1500
Db	1441	ACAATAAGTGTGCTCAGCTCCACATATGTATGTAATAAAACATTCCTCCCGGGGGCGCCATT	1500
QY	1501	caagatctgcagacttaagggcagagacacctaagatgaatgaatgaatttagtg	1560
Db	1501	CAGGATGGCCCACTTAAAGCGAGGAGACAGACTTATAGAGGTAAATGCAATTAATTAAGTG	1560
QY	1561	ggcaaatcccaagaggaagtgttctgcgtlttgagaagcaccagaatctgaaagaaactg	1620
Db	1561	GGCAAAATCCCAAGGAAGTGTGTGCTGTGAGAAAGCCAAAGATGGAAGAACTGTG	1620
QY	1621	agccttctgtctcttcgcaggaagaacgcttccaeccaaggaactgaatgcagagcca	1680
Db	1621	AGCCTTCTGTGCTTCCTTCGCGAGGAAGACGCTTCACCCAAAGGAACCT-----	1667
QY	1681	agccagatgcagattccaaaagaagcagaagcagaagatgagatatgtgttctaacct	1740
Db	1668	-----GAAAGCGAAGAGATGAGATATTTTTCACACT	1701
QY	1741	gattgcaccaggaattcttgacatttgaaagtcaccaatatgatcagagactgcagc	1800
Db	1702	GATGCGACCAAGGAATTTCTGACATTTGAAGTCCCACTTAATGATTCAGGATCTCGAGCG	1761
QY	1801	cttgtgtcagatgtgccaaagtaacccgttcaaaagagaacacgcagatttgggaatctt	1860
Db	1762	CTTGTGTGTCACGTGTCAAAGGTAAACGGGTCAAAAGAAACACAGCAGATTTGGGAATCTTT	1821
QY	1861	gtcaaggtccattatataatgagagagcagactcaaaagatggaaaggtctcgggtgaatgat	1920
Db	1822	GTCAAAGCTCATTTATTTATGTGAGGAGCGACACTCTTAAGATGGAAGGCTTCGGGTGANTAT	1881
QY	1921	caactgataagcagtaaatgtgaaatccctctlttggaagagacaaaccaagatgtccattgaa	1980

1201	acggtgcgaagaagaccccgcatctgaaccaccgcgcctgagcagatagactctaacga	1260	
Qy			
Db	1201	ACGGGCGAGAAAGCACCCCGCATGAAACACCCCGCTGAGCAGATTAAGCTTCACGAAGA	1260
Qy	1261	ctaccctcaatgcacacacccctcgggaaaacacacatccgcctccagccctcgcacctag	1320
Db	1261	CTACCTCATATAGCGCACACCCCTCGGGAAAAACACATCCGCTCCAGCTCGGCACCTCAG	1320
Qy	1321	aatgtaattagtaagactgttaagcagtgtgtttaaacacccaaaaaataaggcaagactc	1380
Db	1321	AATGTAATTTAATAGCAGCTGTAAAGCAGTGGTTATTAACACAAAAAATAAGCAAGAGCGTT	1380
Qy	1381	aatatccagctctaaagaaggtacagaaggtttggatctgaacatctccagatcgaagatga	1440
Db	1381	AATATCCAGCTTAAGAAAAGGTACGAAAGGTTGGGATTCAGCATCCTTCAGAAATGTGA	1440
Qy	1441	acataagtgagctcaagctccaactctatgynaacaactctcccccggggcgcgcat	1500
Db	1441	ACAATAGGTGGCTCAGCTCCCATATGTATGTAATAAATAATCTCCCGGGGGCGGCCATTT	1500
Qy	1501	cagagctgcgcagacttaagcagcagagacagactataagtgaaatgagtagaatttaag	1560
Db	1501	CAGGATGGCGCAGCTTAAGGCGAGGAGACAGACTTATAGAGTTAAATGAGATTAATTAAGTG	1560
Qy	1561	ggcaaatcccaagaagaagtgttctgcgtgtgagaagcacaagaatggaagaactg	1620
Db	1561	GGCAAAATCCCAAGAGGAAGTGTTCCTGCTTGAGAAGACCAAGATGGAAGAACTGTG	1620
Qy	1621	agccctctgcgtccttcgcacgaagaagccttcaccaccaaggaagtgaatgcagagca	1680
Db	1621	AGCCTTCTGCGTCTTTCGCCGAGGAAGACGCTTCCACCACCAAGGGAAGCT-----	1667
Qy	1681	agccagatgacagatccaaaaaagacgaagaacgaagaatgaagataatgttcttiacact	1740
Db	1681	-----GAAAGCAGAAAGATGAGGATATTTGTTCTTACACTT	1701
Qy	1741	gattgcaccaggaattctgacatttgaagtcacacttaagcttaagatcagagcttcagc	1800
Db	1702	GATGGCACCAGGGAATTTCTGACATTTGAAGTCCCACTTAATGATTCAGGATCTCGAGGC	1761
Qy	1801	cttggctgcagtgltcaaaagtaaacggtlcaaaagagaacacgcagatltggaaactctt	1860
Db	1762	CTTGCTGTCACTGTCAAAAGGTAAACCGGTCAAAAGAAACACAGCAGATTTGGGAACTCTTT	1821
Qy	1861	gtcaagctccattataatgtagagagcgacttcaagaatggaagcttcgggtgaatgat	1920
Db	1822	GTCAAGGTCCATTTATTAATGGAGGAGCGACACTCTTAAGATGGAAGGCTTGGGTGAAATAT	1881
Qy	1921	caactgataagcagtaaatgtgaaatccctgttggcgaagacaaaccaagatgtccatggaa	1980

[illegible][illegible]



```
QY 961 aatgatgcatlgtlccagatlaaataatgagccctcgaaatagaagatltgaaacaaga 1020
Db 829 AATGATGTCATGTCAGGATTTAATATGATGCGACCTTCCAAATAGAAATTTGAACAGCA 888
QY 1021 caacataitgttccgccaagcaatgcgtacaccatcatlgtgttcagtgtgttctgca 1080
Db 889 CAACATATATTTTCGCCAAGGCCATGCGTACACCCATCATTTTGGTTCATGTGGTTCCTGCA 948
QY 1081 gcaaataaagacagatagaaacaatcatccaaagtgtgagaagaatlaactaattcaagc 1140
Db 949 GCAAAATTAAGAGAGATATGAACAACATCCAAAGTGAAGAAGCAATTAATCTATCTACAGC 1008
QY 1141 cgtttagcccttgagcagccagatatattgacaacagagatgtgaaagatgcagagctaac 1200
Db 1009 CGTTTACCCCTTGACAGCCAGATATATTTGACACAGAGTGTGAACAGTCCAGAGGCTTCAC 1068
QY 1201 agcgtgcaagagacaccccgacagcaaccacccgctgacagatagactctcaactaaga 1260
Db 1069 ACGGTGACAGAGAGACCCCGACTGAAACACCCGCTGACAGACATAGACTCTCCTCAAGA 1128
QY 1261 ctacctatagcgcaacccctcgagaaacaaccaatccgctccagcctcgcaactcag 1320
Db 1129 CTACCTCATAGCCGACACACCCCTCGGGAAAAACCAACCATCCGCTCCAGCCTCGGCACTCAG 1188
QY 1321 aatgtatttagtaagacgtgtaagcagtggtltaaacacaaataatgagcaagagct 1380
Db 1189 AATGTATTTAGTACGACTGTAGCAGTGGTTATATACACCAAAATAATAGCAAGAGCTT 1248
QY 1381 aatatacagacttaagaagaagatcagaaggttggatcagacatcacttccagaagatga 1440
Db 1249 AATATTCACACTTAAGAAAGATACAGAAAGTTGGATTCAGCATCACTTCACAGATGTA 1308
QY 1441 acaatagtggtctcagctcccaatcatalgtgaaacaatctcccccggggggcgccact 1500
Db 1309 ACAATAGGTGGCTGAGCTCAATCTATGTGAAAAACAATCTCCCGGGGGGGGCACTT 1368
QY 1501 caagatggccgacttaagaagcgagagacagactatagagtaaaatgagtagatcagtg 1560
Db 1369 CAGGATGGCCGACTTAAGCGAGCGAGACAGACTTATAGAGTAAATGAGATTTAGTG 1428
QY 1561 ggcgaatcccaagaagaagtgttgcgtgtltagaagacacaaagtgaagaagtgtg 1620
Db 1429 GGCAAATCCCAAGAGGAAGTGTGTTGCTGTGAGAAAGCAACAAATGGAAAGCACTGTG 1488
QY 1621 agccttctgtcttcgcccagagagacgcttccacccaaggaaactgaaatgcagagca 1680
Db 1489 AGCCTTCTGCTTTGCCAGGAAGACGCTTCACCCAAAGGGAAT----- 1535
QY 1681 agccagatgcgaattccaaaagaagaagaagaatgaggaatgttcttaacact 1740
Db 1536 -----GAAAGCAGAAAGATGAGATATTTGTTCTTACACCT 1599
QY 1741 gatgacacaggaatttctgacatttgaagtcccaacttaagtatcaggatctgcagcg 1800
Db 1570 GATGGACCAAGGAAATTTCTGACATTTAAGTCCCACTTAATGATGATCGAGATCGCAGGC 1629
QY 1801 cttgtgtcagtgtaaaaggttaacccggtcaaaagaagaacaacgagattgggaatcttt 1860
Db 1630 CTTGCTGTCAAGTGTCAAAAGGTACCGGTCAAAAGAAACCAACCCAGATTTGGAAATCTTT 1689
QY 1861 gtcaagttccatttaatagtgagagcagcatctaaagatggaagcttcgggttgaatgat 1920
Db 1690 GTCAAGTCCATTATTTAATGAGAGAGCAGCTCTTAAGATGGAAGGCTTGGGGTGAATGAT 1749
QY 1921 caactgataagaataatggaagaatccctgttggcaagaacaacaagaatccatgtgaa 1980
Db 1750 CAACTGATAGCAGTAATTTGAGAGATCCCTGTTGGCAAGAACAAACCAAGATCCCATGGA 1809
QY 1981 accctaaagaagttcattgtctactggaagcaataacgaggaatgctccagcttaattgt 2040
Db 1810 ACCCTAAGAGTCTATGTCTACTGAAAGCAATTAACGAGAGATGATCAGCTTATTTGT 1869

QY 2041 gcaagagaaataaagaagatgcaatgagctgaagttcaacctggaagccccctggaactag 2100
Db 1870 GCAAGAGAAATAAACCAAGTGCATAGAGCTGAAATCTGACCTGGAGAGCCCCCTGAGCTGAG 1929
QY 2101 ctgcccattgaaacagcgttggatgataagaacgaagaatcccatccctcaagt 2160
Db 1930 CTGCCATTTAAACCAAGCGTTGATGATAGAGAAAGAAATTTCCCATTTCCCTTACAGT 1989
QY 2161 gggatgaggggcttgatgaatcgccacagcaaatgctgcctcagtaagataaggt 2220
Db 1990 GGGATGAGGGGCTTGTATGATGCCAAAGCAAAATGCTGCCCTCAGTAGGATGATG--- 2046
QY 2221 gagtcaggttaataacagacttcccttaagtgaaatgcccagaagtgaactgttaatt 2280
Db 2047 -----GATAAATACCAAGCTGTCCCTACAGTAATATGCCCAAGATGACACTGTATT 2100
QY 2281 atagaagatgacaagtttgcagctgtcttcccaaatctctctgacagctcctccagc 2340
Db 2101 ATAGAAGATGACAGGTTGCAAGTGTCTCTCCACATCTCTGTACACAGTCTCTTCAGC 2160
QY 2341 tcccatgataatgttggttgtgtgacggcagatgtctgacttgggccaaggtgtaatc 2400
Db 2161 TCCCATGATGATGTGGGTTTGTGACGGCAGATGCTGTACTTGGGCCAAAGGCTGCAATC 2220
QY 2401 agtataagccagactgctctttagtcagatgltatcagttccttcttcaaga 2460
Db 2221 AGTATTCAGCCGACCTGCTTTGAGTCCAGATGTTGATTCAGATTCCTTTCACGA 2280
QY 2461 gaagatttggagctgagatgatagtcagaanaacgcacaagaatlttcagatgcag 2520
Db 2281 GAAGATTGTGACGCTCAG----- 2298
QY 2521 caattgatttcgtlaaanaacgaaatacaaaagatgattgatatgctgacag 2580
Db 2299 -----ATAGCTGACGAG 2310
QY 2581 actaaactaatacagtgatgataccagaagaagttctccagcagagatgtgtgtct 2640
Db 2311 ACTAAATCAATATCACTGGATGATGACAGAAAGCAGGTTCTCCACAGACAGATGTGGTCT 2370
QY 2641 tccctgtgtctgaagaagttcaagctcgttggagagttcgcagaccgcaagtgcagagt 2700
Db 2371 TCCCTGTGCTGTAAGAAAGCAAGCTCGTTGAGAGTCTGACGACCGAGTGTGCCGAGGTG 2430
QY 2701 acttgaatgggatatcccttccatctgtccacgycgcggtgaaataacagagcagggga 2760
Db 2431 ACTTGAATGGGATATTCCTTTCCATCTGTCACGCGCGGATTAATCAGAGCCAGGGGA 2490
QY 2761 tgcattgagagcttcaagactgcccacagacaatctatgataaaccgcggtatgat 2820
Db 2491 TGCATTTGAGACTTTCAGAGCTGCCATCGACAAATCTTATGATTAACCCCGGTATGAT 2550
QY 2821 gatgataagcagatgagacttggaaagaacacagaagaagtcaagatcagagggga 2880
Db 2551 GATGATGAGGGCATGGAGACCTTGGAAAGAACACAGAAAGTAAGTTCAGAGTCAAGGAGGA 2610
QY 2881 gagtctgtatccacagcagtgatcagccttccactctctgtgaagacaaatgaaatgga 2940
Db 2611 GAGTCTGTATCCACAGCGAGTGTATCAGCTTCCCACTCTCTGGAGAGCAATATGATGGA 2670
QY 2941 aaccagaagaagtgataagactgataaanaaagaataaactggaagaagaagaagaag 3000
Db 2671 AACCAAGAGAAAGCTATTAAGACTGTATGAAAAAAGGATTAACCTGGAAGAAAAAAG 2730
QY 3001 aaagataagataaagagaagataaanaaagaagaagaagaagaagaagtgcgaaggttg 3060
Db 2731 AAAGATAGAGATTAAGAGAGATTAATTAAGAACCAAGAAAGGGAATGTGAAGGCTGTG 2790
QY 3061 ggaagacatltcaggttggcaaaacatccgaaagaatgacaagaattgagaanaacgggttaa 3120
Db 2791 GGAGACATGTTCAAGTGTGGCAAAACATCGAAAGATACCAAGATTCAGAAAGCGGTAAA 2850
QY 3121 ataanaatacaggaatcccttatacagaaagagagagatgaatgaaagcagagagcag 3180
```



D	b	1009	CGTTTATGACCCCTGACAGCCAGTATATTTGACAAACAGAGGTGTGAACAGATGCGAGGGCTTAC	1068
Q	y	1201	acggtgaaggaagacaccccgacttgaacacacccgccttgaagacagatgaacttcaacga	1260
D	b	1069	ACGGTGCAGAGACACCCCGACTGTGAACACCCCGCTGAGCAGATATACCTCTACCTCAAGA	1128
Q	y	1261	ctactcctaagcgacacacccctcggaaaacccactcgccttcagcctcggaactcag	1320
D	b	1129	CTACCCATAGGCGCACACCCCTCGGAAAAACACATCCGTCCAGCTCGGCACTCAG	1188
Q	y	1321	aatgatatgaagagactgtaagcagtgtgtatatacaccacaaataagagaaagctt	1380
D	b	1189	AATGTATTTAGTACGACTGTGAAGAGGCTTATTAACACCAAAAAATATGCGAAAGAGCTT	1248
Q	y	1381	aatcatcagcttaagaaaggttacgaaggttttggattcagcatcaacttccaagatgta	1440
D	b	1249	AATATCCAGCTTAAAGAAAGGTACAGAAAGTTTGGATTACATCATCTCCAGAGATGA	1308
Q	y	1441	acaataagttgctcagctcccaactctatgtgaanaaacattccccgggggggcatt	1500
D	b	1309	ACAAATAGTGGCTCAGCTCCCAATCATGTGAAAAACATTTCCCCCGGGGGGGCGCAAT	1368
Q	y	1501	caggatgcccagcttaaggcaggaagacagactttagaggttaaltgagtagattagtg	1560
D	b	1369	CAGATGCGCGAGCTTAAGCAGACAGACAGACTTATAGAGGTAATGGAGTAGATTAGG	1428
Q	y	1561	ggcaatccccaagaaggaagtgttctgcgtgttgaagacacaaatggaaggaacttg	1620
D	b	1429	GGCAAAATCCCAAGAGAAAGTGTGTGCTGTGTAAGAGCCACCAAGATGGAAGAACTGG	1488
Q	y	1621	agcctctcgtcttcctgcgcaggaagacgccttcaccacaaagggaactgaatgcagagca	1680
D	b	1489	AGCCTCTTGCTGTTCCTGCGCAGGAAGAGCGCTTCACCCCAAGGAACCT-----	1535
Q	y	1681	agccagatgcagattccaaagaagaagacgaagatgtagaattgtcttacaact	1740
D	b	1536	-----GAAAGCGAAGATGAGAGATATTGTTCTTACACT	1569
Q	y	1741	gattgaccacaaggaaattctctgacatttgaagtcaccacttaatgaattcaagatctgaagc	1800
D	b	1570	GATTGCACCAAGGAATTTCTGCATTTGAAGTCCCACTTAATGATTCAGATCTGCAGGCG	1629
Q	y	1801	cttgggtgcagtgccaaggtaacgggtcaaaagaagaaacacgagaaatttgggaacttt	1860
D	b	1630	CTTGGTGTCAGTGCAAAGGTAAACGGGTAAAGAAACCGCAGATTGGGAATCTTT	1689
Q	y	1861	gtcaagtcctatttaaaggaggaagcaacatcaaatgatgaaggttcgggtgaatgat	1920
D	b	1690	GTCAAGTCCATTATTAATGAGAGACACACATCTPAAGATGGAAGGCTTGGGTGATGAT	1749
Q	y	1921	caactgatagcagtaatggaagaatccctgttggcaagaacaaacaaatgcatctggaa	1980
D	b	1750	CAACTGATATGCAGTAAATGGAATCCCTGTGTGGCAAGCAACCAACATGTCATGGA	1809
Q	y	1981	accctaaagaagttctatgtctactgaaaggaataaaccggaatgatccaagttatgtt	2040
D	b	1810	ACCTTAAGAGAGGTCTATGTCTATCTGAAGGCAATTAAGAGGAATATCTCACTTATTTGT	1869
Q	y	2041	gcaaggaagaataagcaagtgccaatgaaacttgaatccctgggaagcccccctggaagc	2100
D	b	1870	GCAAGGAATAATACCAAGTGCAATGAGCTGAATCCCTGTGGAGAGCCCCCTGAGCTGAG	1929
Q	y	2101	ctggccatttgaagaacggttgatgatatagaagcaagaatttccattcccttcaagt	2160
D	b	1930	CTGGCCATTGAAGAAAGCGTTGATGTATGAAGAAACGAATAATTTCCATTCCTCTACAT	1989
Q	y	2161	ggagattgaagggcttgatgaatgcgcgcagcagaatctgcctcagtagaataatgggt	2220
D	b	1990	GGGATTGAGGGGCTTGATGAATCCCCACGCGAATGCTGCCCTCAGTAGAATAAG---	2046
Q	y	2221	gagtcaggttaataaccagctgtccctacaatgataatgccccaaataaactgcatc	2280
D	b	2047	-----GTTAAATACCAAGCTGTCCCTCTACATGTAATATGCCCCAAATATACACTGTCAATT	2100

QY 2281 atagaagatgacaggtctgcccagctctccacacatctctgaccagctcctctccagc 2340  
 |||||  
 DB 2101 ATAGAGATGACAGCTCCAGTCCCTCCACATCTCTGACACAGCTCTCTCCAGC 2160  
 QY 2341 tcccctgctgctggtggttctgacgacagctgctgctgctgacgacagctcctc 2400  
 |||||  
 DB 2161 TCCCATGATGATGAGGCTGTTGTGACGGCAGATGCTGTGACTTGGCCCAAGCTCCAAATC 2220  
 QY 2401 agtgcctcagccagctgctcttgctcagatgctgacatgctgacatgctctcttcacga 2460  
 |||||  
 DB 2221 AGTGATTCAGCCGACTGCTCTTTGAGTCCAGATGTTGATCCAGTCTCTTCTTTCACACA 2280  
 QY 2461 gaaagatttgagcgtcagagatgctcagaaaaacgacaaaagcaalttcagatgcccag 2520  
 |||||  
 DB 2281 GAAGATTTGGACGTACAGATGTCAGAAAAACACAAACCAATTTTCATGCTCCAGT 2340  
 QY 2521 caattgagatttcgttaaaaaacagaaaaatcaaaaaacatgagatttgatagctgaag 2580  
 |||||  
 DB 2341 CAATTGATTTGTTTAAACACGAAAAATCAAAAACATGGATTGATGTA----- 2389  
 QY 2581 actaaactcaatacagtgatgacagaaagcaggtctctccagcagagatgtggtcct 2640  
 |||||  
 DB 2390 -----GTTCTCCACGACGAGATGTGGTCT 2415  
 QY 2641 tccctggtgctgaaagaaagcaagctcgttgagagagctgacagccgagctgcccaggtg 2700  
 |||||  
 DB 2416 TCCCTGGGCTGGAAGACGACGCTGTTGAGAGCTCTCAGACGCGCAGATGTTCCGAGGTG 2475  
 QY 2701 acttgagagagagatct 2760  
 |||||  
 DB 2476 ACTTTGAATGGGAGATTTCTCTTCCATCTCTCCACGCGCCGCGATTAATCCAGAGAGGGA 2535  
 QY 2761 tgcagtgagagctcagagctgacatgcacaaatcttatgataaaccccggtgagatg 2820  
 |||||  
 DB 2536 TGCATGAGAGCTTCAGAGCTCCATCGCAATCTTATGATTAACCCCGGTAGATGAT 2595  
 QY 2821 gatgatgaagcagctgagagaccttggaagaagacacagaaagaaagctcaagatcagggaga 2880  
 |||||  
 DB 2596 GATGATGGAAGGATGAGACCTTTGGAAGAAACACAGAAAGAAATGTCAGAGGAGA 2655  
 QY 2881 gagctgctgacagagcagtgatgacgctctccacactctctctgagagagaaatgaa 2940  
 |||||  
 DB 2656 GAGTCTGATTCACAGCCAGTATGATCCACTCTCTGAGAGCAAAATGAAATGGA 2715  
 QY 2941 aaccagagaaaggtgataagactgataagaagaagaagataaaactgaaagaagaagaag 3000  
 |||||  
 DB 2716 AACCAAGAAAGGTGATAGACTGATGAAAAAAGATTAACCTGGAAGAAAAAGAAAG 2775  
 QY 3001 aagaatagaagataagagaagaataaaatgaaagccaagaaggaatgctgaaaggtctg 3060  
 |||||  
 DB 2776 AAGATATAGAGATTAAGAGAGAGATTAATAATGAAGCCAAAGAGAGATGCTGAAGGCTTG 2835  
 QY 3061 ggaagacatgctcaggttctgcaaacatcgaaaaagatgacaagaattgagaagaaggt 3120  
 |||||  
 DB 2836 GAGACATGTTCT----- 2847  
 QY 3121 ataaaaatacagaatccttaccatcagaagaagagagatacgaatgaagcagagcag 3180  
 |||||  
 DB 2848 ----- 2847  
 QY 3181 gagaagatccaagccaaaactcgagaatttgaaggaaagcagagctcgagagcgtgactat 3240  
 |||||  
 DB 2848 ---AGATTTCAAGCCAAACTCGAGAAATTAGGAGAACGACAGCTCGAGAGCGTGACTAT 2904  
 QY 3241 gctgaagaatccaagatttccatcgagacatttgctgctgctgctgctgctgctgctgctg 3300  
 |||||  
 DB 2905 GCTGAATATTCAGATTTTATCGGACATTTGCTGTGATGATGATGATGATGATGATGATG 2964  
 QY 3301 gttctctctatgaaggttcacatgctgctcgaagctcagagctcagagcccaagagaag 3360  
 |||||  
 DB 2965 GTTCTCTTATGAAGGTTCATGCTCTCAACGCTAGACCTCAGAGCCACAGAGAAAGG 3024

QY 3361 catatgatgagatgcttctgctgctgctgctgctgctgctgctgctgctgctgctgct 3420  
 |||||  
 DB 3025 CATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3084  
 QY 3421 gtgagacagtaacagatacactcctctagcaatcctgctgctgctgctgctgctgctgctg 3480  
 |||||  
 DB 3085 GTAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3144  
 QY 3481 ttccagcaagcaaaagaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg 3540  
 |||||  
 DB 3145 TTTTCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 3204  
 QY 3541 caacctggtcgaaagcagc 3600  
 |||||  
 DB 3205 CAACCTGCGCCCAACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3264  
 QY 3601 gtgcagatgcagc 3660  
 |||||  
 DB 3265 GTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3324  
 QY 3661 taacgctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 3720  
 |||||  
 DB 3325 TACAGCTCTCTGCTCGCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 3384  
 QY 3721 gaggcagaactactcctcctggtggtggtggtggtggtggtggtggtggtggtggtggt 3780  
 |||||  
 DB 3385 GAGCAAACTACTCTCCCTCGGAGAGGCTTCCAGAGTCCAAAGAGAAACCCAGGTACTCTC 3444  
 QY 3781 agctaccagagctcag 3840  
 |||||  
 DB 3445 AGCTTACCAAGGCTCCAGGAAAGGCTTACTGAGAGCAATGAGCTTCAACGCCAGAGGCTAG 3504  
 QY 3841 ctggaactcag 3900  
 |||||  
 DB 3505 CTGGAATCTCAGAGAGCTCTTCCGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3564  
 QY 3901 cagctcctcctcag 3960  
 |||||  
 DB 3565 CAGCCCTCTTCCGAGAGGCGCCAGCAACTATGACTCTTATTAAGAAATGTCAGAGACCCCACT 3624  
 QY 3961 taagccctcctcag 4020  
 |||||  
 DB 3625 TACGCCCTCTCCAAAGGAGGCTTCCGCAAGATGATGATGATGATGATGATGATGATGATG 3684  
 QY 4021 aggtctgaacagactcagactcctgagaaagagagagagagagagagagagagagag 4071  
 |||||  
 DB 3685 AGGCTCAACAGACTTCAGACTCTGAGAAAGGAGGCGCTTATTCCTCGA 3735  
 |||||  
 DB 3735 AGGCTCAACAGACTTCAGACTCTGAGAAAGGAGGCGCTTATTCCTCGA 3735

RESULT 9  
 AB005549  
 LOCUS 5500 bp mRNA linear ROD 13-NOV-1998  
 DEFINITION Rattus norvegicus mRNA for atypical PKC specific binding protein,  
 complete cds.  
 ACCESSION AB005549  
 VERSION AB005549.1 GI:3668777  
 KEYWORDS asbp; atypical PKC specific binding protein.  
 SOURCE Rattus norvegicus cell\_line:3Y1 cDNA to mRNA.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 5500)  
 REFERENCE 1 (bases 1 to 5500)  
 Izumi, Y.  
 DIRECT SUBMISSION  
 JOURNAL Submitted (03-JUL-1997) Yasushi Izumi, Yokohama City University  
 School of Medicine, Molecular Biology, 3-9, Fuku-ura, Kanazawa-ku,  
 Yokohama, Kanagawa 236, Japan (E-mail:izumi@med.yokohama-cu.ac.jp,  
 Tel:045-787-2597, Fax:045-785-4140)  
 2 (sites)  
 REFERENCE  
 Izumi, Y., Hirose, T., Tamai, Y., Hirai, S., Nagashima, Y., Fujimoto, T.,  
 Tabuse, Y., Kempfner, K.J. and Ohno, S.  
 An atypical PKC directly associates and colocalizes at the



epithelial tight junction with ASIP, a mammalian homologue of  
*Caenorhabditis elegans* polarity protein PAR-3  
 JOURNAL Cell Biol. 143 (1), 95-106 (1998)  
 MEDLINE 98437350

FEATURES  
 SOURCE location/Qualifiers

1..5500  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /cell\_line="3Y1"  
 260..4273  
 /gene="asbp"  
 260..4273  
 /gene="asbp"  
 /codon\_start=1  
 /product="atypical PKC specific binding protein"  
 /protein\_id="BAA34216.1"  
 /db\_xref="GI:3868778"  
 /translation="MKYVCGFRIRRVVPCGGDKMKFSLIQAVTRKAVAKDPY  
 WIVYRLHEDGGILDDLDLIDVADKDLVAVFDQDPHGGDGTASSTGQSE  
 IFSELGNTNVSAFRPYQTTSELEVPSVIRANMPLHVRSSDPALGLSTVDNF  
 SSEPSKRNPTRMSTAGFLKQNTGSPKCDKDKENDYKSLPDPSSMNOFORNA  
 RSSLSAPHVMDRLKEKODEDEGTEDSSRVEPVGAADGLEMPFSLDMDKLVQ  
 VPDGGLGIHVVPESARGRTGLVYKLEKGAQENMLFHNOCIVRLINDGLN  
 RREQAHMFRQAMRAVIRVHYVPAANKQYEDLSRENNNTSPCFSPDSHCVANR  
 SVANNAQALPRAPRLSQPEQDLDAEPLRPHSAHAKTPPTAPALPVALDSTVGSV  
 YNTRVGRKRLNIDLRKTEGLGFSITSRDVYIGSAPVYKNIIPRGAAIDQGRKAG  
 DRLLEVGVOLAKSOEEVYSLRSTKMEIVSLVROEAEPREMADEPOMOSP  
 KETRAEDEDIVLPDGRFELTFEVLNDGSAAGLVAGVSGNSKEMHADLGEVKS  
 INGASAKDRRLRVNDOLIVNGESLIGNOEAMETVLRSMSEKGRMIDOLIVR  
 RISRNEILRSPGAPAEPLIETELDRLEPRLSHSYSGIEGLDESTRNAALSRING  
 ESKCOLSPYVNMHDDTVMIEDRLPEPLHSDOSSSHDVGITMEAGWAA  
 TISDSACSLSPVDVPLAFVOREGFGQSKSEKRTKQFNASOLDPEVTKRSKMDLG  
 IADETKLNTVDVDRAGSPNDVPSLGLKSSLESLOTVAETLLNGNTPRPR  
 IIEGRGNESEFRAIDKSYDKPWMDDEEMETLEEDTESSSGRESVTSDDQPE  
 SLEROMNGDEPEKDKAEKKDKAKDKKKREKDKLKGKLGKIGMTPFRKIR  
 KDKMEKGRKIKIDQSTSEDRVMEQERIOAKREREREOARVDVATODPFR  
 TRGCDDELYAGSSVYDGLAANARQSPREGLMNTLYQVKKPKRSKGDNRSTP  
 SNDRIORLROEFOAKODEVEDRRTITSEDSMSRPAOSGRHSVSEVVOVKQ  
 ROBERSFOAOROYSSLPQSRKMSVSDSEMDQYARGEFOSKENDRPSYOG  
 SRNGYLGHGFNARVMLETOELRROERREKQDLKQPPADVGVRGPRQDPVPSPOY  
 ARNLRLQTPKRGAPFYS"

BASE COUNT 1450 a 1449 c 1547 g 1054 t  
 ORIGIN

Query Match 53.7%; Score 3056.4; DB 10; Length 5500;  
 Best Local Similarity 79.5%; Pred. No. 0;  
 Matches 3803; Conservative 0; Mismatches 856; Indels 123; Gaps 11;

QY 1 atgaagtgacgctgtgtcttcgaagcagccgggtgtgtcgtgcggtgcggagcagccac 60  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 260 ATGAAATGACCGTGTCTTCGGAGAGACCGGGTGTGCGGTGCGGAGATGGCGCG 319  
 QY 61 atgaagttttcaagcctcaccagcagcggtgacccgctaccaggaagcgaatgccaag 120  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 320 ATGAAAGTTTTCACGCTTATTCACAGCAGCGGTGACCGCTACCGGAAGCCGTGCCAAG 379  
 QY 121 gatacaactactgagatagcaggtgcatcgcttggaacatgagagatggaataataagac 180  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 380 GATCAACAATCTAGTATGATACAGGTACATCGTTGGAGCATGTGGAGCGGCGGATCTTGAC 439  
 QY 181 ctgtatgacattcttctgtatgtagcagaagataaagacagactgtgtagcagtttggat 240  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 440 CTGATGACATCTCTGTGACGTTGCGATGACMAAGACAGACTGTGACAGTGTGGAT 499  
 QY 241 gagcagatcacatcacagcaggtgtgtagcagctgacagttccagagttaccacaagc 300  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 500 GAGCAGATCCCCACCAATGAGAGAGATGTACGATGCCCACTCCAGGGGAACCAAGT 559  
 QY 301 ccagagataattgtagttagcagcttgaccacaacatgtctcagccttcagccttaccaa 360  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 560 CCAAGATATTTCGGCAGTGGAGCTGGCACCACCAATGTTTCTGCTTTCGGCCTTATCAA 619

QY 361 gcaacaagtgaattgagatcacaccttcagcttccttcgagcaaatatgctcttcattgt 420  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 620 ACCCAATGAAATATGAGTACACGCTTCAGTCTTCTGGGCAAAATATCTCTTCACGTC 679  
 QY 421 cgaagcagtagttagcagcagctctaatgttgccttcaccctctctgtcagtagtaattt 480  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 680 CGACGAGACAGTGAACCCAGCTTTAACTGGCCCTATCCACTCTCGTACAGATAAACAACCT 739  
 QY 481 tcccttgagagccttcgaagaaaataaccacacgctgtgtataaacaagcctgtcctc 540  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 740 TCTTCAGAGGAGCCCTCCAGAAAACCCCAACCGCTGTCTCACAGCAGCTGGCTTCTC 799  
 QY 541 aagcagaacactgctgagagctcctaaacctgcagcaggaagaaagatgaataactacaga 600  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 800 AAGCAACACCACTGGAAATGCCAAAACCTGCGCAGAGAAAGAAATGAATACTACAGG 859  
 QY 601 agctcccgaggatatactagttaactgtgtcacaattcagaagagacatgtctcgtcg 660  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 860 AGCTCCCAAGGATCCAGTAGCTGTGCCAACAGTTCCAGAGAGACAAACGCCCGCTCC 919  
 QY 661 tctctgagtgacgtacaccatgtgtggcagagtgtgtggaagaaacagaacagatgag 720  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 920 TCTGTAGTGGCCACCAACCAATGTTAGACCGGTGCTGAGAAACAAAGACAGAGAG 979  
 QY 721 gatggacaagaagagataacagtcgtgttgaaacctgttgacatgctgacacaggtttg 780  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 980 GAAGCACAAGAAAGACAGACGCGAGTGAACCGGTTGGACATGCTGATACAGAGACTG 1039  
 QY 781 gagcatatacccaacttctctctgtgagatagatagttaaagctcgtagaagttcccaagat 840  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1040 GAGATATGCCCAACTTTCTCTCGATGATATGTTAAAGCTGTCAAGTCCCAACAGAT 1099  
 QY 841 ggaagagccttcgggaatccatgtatgtccttcagcttcgagcggcagaacaccttggg 900  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1100 GGAAGGCCCTCGGGAATCCATGTTAGTCTTCACTGCTGACGGCGCAGAACACTGGGG 1159  
 QY 901 ttattagtaaaacgattggagaaggtgtgtaaacgtcgaacatgaanaactcttctgtgag 960  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1160 TTGTTATGAAAGCGGTTTGAGAAAGGTGGAGAGGCTGAACMAAACCTTTTCCACAGAG 1219  
 QY 961 aatattgcattgtcaggaattaatgattgagcagcttcgaaatagaagaatttgaacaagca 1020  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1220 AATGACTGCAATGTGAGAGATTAAAGATGAGATCTTGGAAACAGAAATTTGACCAAGCA 1279  
 QY 1021 caacatatttgcgaagccatcgtacacacatatttgcttcattgttcttcga 1080  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1280 CAGCATATGTTCCGCAAGCATGCTGCGGATCTATTGTGTTCCACGTGCTCCTGCA 1339  
 QY 1081 gcaaatagaagcagtatgaaacaactatcccaaggtgagaagaacaaattactattcaagc 1140  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1340 GCGAACAAGGAGAGTATGAACACTCTCCACAGAGAGATGAACAACTATTCCTGCGC 1399  
 QY 1141 cgtttgacctgacgacagatattatgacaacagaggtgtgacaatgtcagagcttcac 1200  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1400 CGCTTCAGCCCGACACCACTGTGTGGCCAAACGAGGTGTGGCCAAACATGCCCTCAG 1459  
 QY 1201 acggtgcagaagcaccgcagataaccacccgcttgaagcagatagactctcaataaga 1260  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1460 GCGTTGGCCAGAGACCAAGACTGAGCAGCAGCACTGAGCAGCTGTGATGCTCACCCCGA 1519  
 QY 1261 ctactcatagcagacacccctcggaanaaccacatccgctcagcagctcgacactcag 1320  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1520 CTACTCATATGTTGCTCATGCTCAACCAACCCGCCACAGCCCAAGCTGTGCTCCACCG 1579  
 QY 1321 aatgtattgtagtactgtaagcagtggttataacacccaaataaagcaagaagctt 1380  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1580 AATGTGCTTATGATCAATGTGTGCGAGTGTGTACAAACAGAAAGATGAGCAAGAGCTC 1639  
 QY 1381 aatatccagtttaagaaagttacagaaggttggagtttcagcatcattccaagatgata 1440  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 1640 AATATTAGCTTAAAAAAGGTACGAAGAGACTGGGATTTTGCATCTTCCAGACAGTA 1699  
 QY 1441 acaataggtgtcagctccaatctatgtgaaaaaatctctccccgggggagggcatt 1500



```
Db 1700 ACTATAGGAGGCTCAGCTCCATTATATGTCAGAAACATTCTCTCGAGGGGCTGCATT 1759
QY 1501 caagatggccgaagtaagcaagagacacactatagaagtaagtagatagttagtt 1560
Db 1760 CAGGATGGCAGGCTCAAGCGGGAGACCGCGCTAATATGAGGTCAATGAGTAAATTAGG 1819
QY 1561 ggcacaccccaagaggaagtgttttcgtgtgtgtaagacacccaagaatggaagaaatgt 1620
Db 1820 GGCAAGTCCAGAGGAGGAAGTTGTTCCCTGTTGAGAGACCAAGATGAGAGGAGACCGTG 1879
QY 1621 agccttctgttcttcgcagagaagacgcttcacccaaggaagtaatgcagagcca 1680
Db 1880 ACCCTTCTGGCTTTCTGTCAGAAAGAGGCTTTCACCCCAAGGGAATGATCCGCAACC 1939
QY 1681 agccagatgcagattcccaaaagaaacgaagcagaagaatgagatattgttcctaact 1740
Db 1940 AGCCAGATGCCAAGTCCAAAAGAAAGCAAGCCGAAGACGAGACATTGTTCTCACACTT 1999
QY 1741 gatgcacccaaggaattcttcgacatttgaaagtcaccaatgatcaggatctgcaggc 1800
Db 2000 GACGGTACCGAGGAGTATTCTGACTTTCGAAAGTTCACACTGAATGACTCAGGGCTCGAGG 2059
QY 1801 ctgtgtcagtgtaaaaggtacacggtcaaaagaaagacacagcagatttggaatcttt 1860
Db 2060 CTTGGTGTGACGCTCAAGGGGGAACCGCTCCAAGAGAACCATGCAATTTGGGATCTTC 2119
QY 1861 gtcaagctcatattatgatgaaagagcagcatcctaagaatgaaagcgttcggtgtaagt 1920
Db 2120 GTTAAAGTCCATATTAAACGGGGAGCTGCATCTAAAGATGGGAGCGTGAGCGTAAATGAC 2179
QY 1921 caactgataagcaftaatgagaatccctgtctggtgcaagaacaacccaagaatgcattgaa 1980
Db 2180 CAGCTGATAGCTGTATACACGAGAGAGTCTGTGGGCAAGCCAACGAGAACCATGAGAG 2239
QY 1981 accctaagaaggtctatgtctactgtaagcaataaagaggaatgatacagcttatgtt 2040
Db 2240 ACTCTGAGAGGTCATGTCATGTCACCAAGGGCAACAAGCTGGCATGTCAGCTCATCGTG 2299
QY 2041 gcaagagagataatgaagcaatgacatgacatgacatgacatgacatgacatgacatgac 2100
Db 2300 GCGAGGCGGATGACGAGATGTAACGAGTTGCGGTCTCTGGGAGCCCGCTGCACCCGAG 2359
QY 2101 ctgcccatgaaacacgctgtgtgataagagaacgaagaattcccatctccctcactcag 2160
Db 2360 CTGCCATATGAAACAGAGTTGGATGACAGAGAACGCGATCTCACACTCCCTACAGT 2419
QY 2161 gggatctgaggggtctgataatcgcgccagcagaatgctgcctcactgataagttgg 2220
Db 2420 GGGATCGAGGGGCTGATGATGCTCCACACAGGAATGCCGCACTCGACGAGATATGGGT 2479
QY 2221 gagtcaggtaaatcacagctgtccctacacagtgaaatgacccaagaatgacactgtcatt 2280
Db 2480 GAGTCAGGTAAATGCGACAGCTCTCCCAACCGTGAACATCCCAATGATATACACTGTCAAG 2539
QY 2281 atagaagaatgacagatgtccagtgcttcctcacatctctctgtgacacagtcctctcag 2340
Db 2540 ATTGAAGATGACAGGCTGCTGTGCTCTCCCTCACTCTGACCAAGTCTCTCTCAACG 2599
QY 2341 tcccatgatgatgtgggtgtgtgacgagagatgctgtaactgtggcgaagctgtgaatc 2400
Db 2600 TCCCATGATGAGGTGGGATTCATTAATGACAGAAAGGACGCGGGCCAAAGGCTATCCATC 2659
QY 2401 agtgaattcagccagactgtcctctgtgacagatgtgtgacacagattctgtcttcaacga 2460
Db 2660 ACTGACTCAGCGGAGTGCATTCATTTGAGTCCAGATGTTGATCCGGTTCCTTTTCAACGA 2719
QY 2461 gaaggaatttgacgctcagatgtatgtcagaanaacgacaaagaatcttcagatgcag 2520
Db 2720 GAAGGATTGAGCGCGAGTATGTCTAGAAAAACGACAAAGCAATTTCAATATGCGAGT 2779
QY 2521 caattggaattgtttaaacacagaataacaaagaatgattgttgatgactgcag 2580
|||||
```

---

```
Db 2780 CAATTGCAATTTGCTTAAACACAGAAATCAAAAACATGATGATTAGCTATGACGAG 2839
QY 2581 actaaactcaatacagltggaatgaccagaagacaggtcttcocagcagagatgtggtcc 2640
Db 2840 ACTAAATCTCAATATCACTGATGATGACGAGAGCAGGCTCCGCCAATATGATATGGGAGCC 2899
QY 2641 tccctggtctcgaagaatcgaagctcgtgtgagagtcgtcgaacgagatctgcagagttg 2700
Db 2900 TCCCTTGGTCTGAGAAATCCAGACTCTTTAGAAAGTCTGACAGCGGCTGTGCTGAGGTG 2959
QY 2701 actttgaatgggagatatttccttcacatgctccagccgcgcagataatcagagcagga 2760
Db 2960 ACCCTGAATGGAACATTCCTTTTCACCGCCACCGCCAGAAATCATCCGAGGAAGGGCC 3019
QY 2761 tgcataagagcttcagagctgcacacagcaaatctcatalgatataaacccgcgtagatgat 2820
Db 3020 TGCAGAGAGAGCTTCAGACCGGCATTTGACAACTCTAGATATAGCCATGTGTGATGAC 3079
QY 2821 gatgatgaagcaatggaagaccttgaaagagacacagaagaagtccaagatcagagagaga 2880
Db 3080 GACGACGAAGGCGCATGGAAGACCTTGAAGAAACACAGAAAGAAAGTTCAAAGTCAAGGAGG 3139
QY 2881 gagctgataccacagccagctgataagccttcacactctctggaagagacaaatgaaatga 2940
Db 3140 GAGTCCGTGTCCAGTCCAGGTGATGACCTTCTTATTTCTGTGAGAGACAAATGAATGGA 3199
QY 2941 aaccagaagaaagtgataagactgaagaaagaaagataaaacttgaaagaaagaaag 3000
Db 3200 GACCCAGAGAGAGGAGCAAAAGCAGAGAAAGAAAGAACCCGGAAGAGATTAAGAG 3259
QY 3001 aaagataagagataaggaagaagataaaatgaaagccaagaagggaattgtcgaaggtctg 3060
Db 3260 AAAGACCGAGAGAGAAAGAGAGATTAACGAAAGCCAAAGAGGAGGTGCTGAAGAGGCTTG 3319
QY 3061 ggaagacatttcaggttttgcaacacatcgaaagaatgaaagattgaagaaaggttaaa 3120
Db 3320 GGGGACATGTTCAAGTTTGGCAAAACATCGAAAGATGACAAAGTGAAGAAATATGGCTCGA 3379
QY 3121 ataaaaatacagaagatcccttaccatcagaagaagagagaataagaaatgaaagcagag 3180
Db 3380 ATAAATATCCAGAGACCTTTTCACTCAGAAAGAGACAGGCTGGGATGAAGAGAGAAACAG 3439
QY 3181 gaagagattccaagccaagaactcgagaatttaggnaacgacgaactcgagaagcgttgat 3240
Db 3440 GAGAGATTCACAAACCAAACTCGAATTTAGGAGACGCGCAACTCGAGAGCGCGACTAT 3499
QY 3241 gctgaatccaagaatttcaatcagacatttgctgtgataagatgataagtaagtaaggg 3300
Db 3500 GCTGAATCCAAAGATTTCCATCGAACAATTTGCTGTGATGACAGATGCTGTATGGGGGC 3559
QY 3301 gttctcttataagagttccatggtctccaagcctlagacactcagaagcccaagagaag 3360
Db 3560 ATGTATCTCATATATGTGGCTGTGGCTCTCAATGCGCAAGCCGAGGCCCAAGAAAGAT 3619
QY 3361 cataatgataagcttctgtatgctccaagatcaagaagccgcggaattccaaacctcaact 3420
Db 3620 CACCTGATGAGACCTTTGATGACCAAGTAAACAACTCGGAGCTCCAA-----ACCT 3673
QY 3421 gtacagatgaacagatcaactctcagcaatcagatcagatcagatcagatcagatcagatc 3480
Db 3674 GGGGACAGCAATGATTCACCCCTTACCAACCATGACCGGATACCACTGTACGGCAAGAA 3733
QY 3481 ttccaagcaagcaaaagataagatgataagatgataagatgataagatgataagatgataag 3540
Db 3734 TTTCACAGACGCAAAACAGAGATGAGAGCGTGGAAAGACCGGGCGCTACCTACACTTTGAG 3793
QY 3541 caacctgtgcggaacgcaagcgcgagcagcagcagcagcagcagcagcagcagcagcagcag 3600
Db 3794 CAATCTTGTGCTCCAGCTCAGCGGCGAGCATCGCAAGATGTGTGCGACTCTGTGATGAGGAG 3853
QY 3601 gtgcacatcagcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3660
Db 3854 GTTCAAGTGCAGCGGACCGCCACGAGAGACGAGAGACCTTCCAGCAGGCCCGCCAG 3913
|||||
```

QY 3661 tacagctctgctcgcaagcaggaagaaatgcaagctcggtctccagagctctg 3720  
 Db 3914 TACAGCTCTGCTCGCAAGCAGGAAGAAATGCCAGCTCTATCAAGGATTCCTGG 3973  
 QY 3721 gaagaaactactccctgggaaagcttcacagatgcaagaagaaaccagatctcc 3780  
 Db 3974 GAAGAAACTACTCCCTGGGAAAGCTTCCAGCTGCGCAAGAGAAACCCAGGATATCC 4033  
 QY 3781 agctacaagagctccagaaagctacactgagagagagagagagagagagagag 3840  
 Db 4034 AGTACAGAGGCTCCAGGAAGCGCTATCAAGTGGCCATGCTTTAAATCCAGGCTCATG 4093  
 QY 3841 ctggaactaagagagctctcgccagagagagagagagagagagagagagagagag 3900  
 Db 4094 CTGGAAGCCAGAGAGCTTCTCCGACAGGAGAGAGCGGAAAAGACAGCACTGAAGAG 4153  
 QY 3901 caagctctctcgagggcgccagcaactatgctctataagaagagctcagagccag 3960  
 Db 4154 CAGCCACAGCGGATGAGTCAAG----- 4177  
 QY 3961 taagccctcccaagggcgcccttcgagagatgccccctccctctcagatgtcg 4020  
 Db 4178 -----GGGCGCTTCCGGCAAGATGATACCCCGCTCCATCAGAGTGCC 4222  
 QY 4021 aagctgaagagagctcagactcctgagaaagagagagagagagagagagagagag 4077  
 Db 4223 AAGCTGAAGAGCTCAGACACCGAGAAAGGCGCCCTTCTACTCTAGGCTCAATG 4282  
 QY 4078 aataacgagatgctcagctgctgcaataaagaacattctcctatgagagctgtat 4137  
 Db 4283 AGATAGGAGAGATTTGCTGATCAATTAAGGACATTTCCCGTGAAGACTGTCTC 4342  
 QY 4138 gggagattttttaaacc-----tcagtgactatgagagatctctgttct 4185  
 Db 4343 TGGAGTTATTTTAAACAAAACCCCTCATTGTGTGACTACAGAACCTTCTGTCACT 4402  
 QY 4186 ggtatcagctgcttggagagctgagcaagaagaatggaagagagagagagagagag 4245  
 Db 4403 GGTATCAGTGCCTTTAAGCAGAGAGCAAAAGAAAGAGAGAGAGAGAGAGAGAGAG 4462  
 QY 4246 tatgtcctaagctgcttctcagagagagagagagagagagagagagagagagag 4305  
 Db 4463 -----GCCTATATCTCGTGAAGAGACCCCATCCCAACCCCAACCTCAAGT 4512  
 QY 4306 ggttcaatagctctgagccctcctcagagagagagagagagagagagagagagag 4365  
 Db 4513 TCAAGGAGATCTACTCCCTCTCTCAGCAGTACCGGAAACTTCACTCATATTAAG 4568  
 QY 4366 gtccgagctgtgttaacatccctgcaacactgagagagagagagagagagagag 4425  
 Db 4569 -----GTGCGTGTGAGAGCTCAAGCTTCACTTCAATTAAGAG 4602  
 QY 4426 aggatgagctgagccctgttctccctcctcctcctcctcctcctcctcctcctc 4483  
 Db 4603 GGGACCCCCCCCCCCCCCAGCGTGTGCGACACCCGTTCCCTGTGAGAGTCACT 4662  
 QY 4484 tagctctgag 4543  
 Db 4663 CGGTCTGTGAGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4722  
 QY 4544 cacaacag 4603  
 Db 4723 CGCAG 4780  
 QY 4604 tcttttgaagc-ggaagagctgcatcctcctcctcctcctcctcctcctcctcctc 4662  
 Db 4781 TCTTCGAG 4840  
 QY 4663 tttagtacc-----atggggagctcttctcctcctcctcctcctcctcctcctc 4718  
 Db 4841 TTtTGtTCTGGGGGTGAGAGGTAATTTCAACTGCTTATTAATCTCAAC--ACCATAT 4898

QY 4719 agtatactctgctgctgtattacatacagctgtcg 4760  
 Db 4899 CACAGGATTTCTAGTCAGTTTATTAAGTACATGCGCTTAGT 4940

## RESULT 10

AF196186 3470 bp mRNA linear PRI 16-NOV-2001  
 LOCUS Homo sapiens atypical PKC isoform-specific interacting protein  
 DEFINITION short variant mRNA, complete cds.  
 AF196186  
 VERSION AF196186.1 GI:13491611  
 KEYWORDS human.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 AUTHORS Fang, C.M. and Xu, Y.H.  
 TITLE Down-regulated expression of atypical PKC-binding domain deleted

JOURNAL Cell Res. 11 (3), 223-229 (2001)  
 MEDLINE 21520080  
 PUBMED 11642408  
 REFERENCE 2 (bases 1 to 3470)  
 AUTHORS Fang, C. and Xu, Y.  
 TITLE Exon/Intron Structure and Splicing Variants of a Novel Human

JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 3470)  
 AUTHORS Fang, C. and Xu, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-Oct-1999) Laboratory of Molecular and Cellular  
 Oncology, Shanghai Institutes for Biological Sciences, Chinese  
 Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, P. R. China

## FEATURES

## source

1. 3470  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /map="10p11.2"  
 288..3383  
 /note="hasIPsa"  
 /codon\_start=1  
 /product="atypical PKC isoform-specific interacting  
 protein short variant"  
 /protein\_id="AAK27892.1"  
 /db\_xref="GI:13491612"

## CDS

/translation="MKVTVCFGRTRVYVPCGDHMKVFLIDQAVTRYRKAKAPNY  
 IIOVHRLEHGDGILIDLDILCDVADKRLVAVEPEDDPHGGGTSASSTGDSPE  
 IFSELEGTNNVSAFOPYOATSELEVPSVLNMPDLYRRSSDPALIGLSTVSQSP  
 SSEPSRKNTKMTSTAGFLKONTAGSPCTCKRDKDENTRSLLPRTSMNSQFQDNA  
 RSSLSASHPVWGKLEKODDEGTEDNSRVEPVGHADTGLEHLPNFSLDVFLVE  
 VPNDGPIGTHVVPARGRTIGLVRLKLEGGKAHEHNLFRENDCIIVRIDNGDILVE  
 RRFEOHMRKQMRKPIIWFHVVPANKREYQLOSEKNNYSRSRSQVITDNR  
 SVNSACLIHVORAPRLNHPPEQIDSHSRLLPHSAHSPGPPSAAPASAPNSVSTVSSG  
 NTNKGKRLNLIOLKRGTEGLGFSITSRDVPTIGSAPLYVKNITPRGAIODGRKAG  
 DRILEVNGVDLVKQSEEVVSLRSTKMGRTYSLVPRQEDAFHREINAPSOQIP  
 KETKADEDIVLPDGTREFLEFEPINDSGAGIGVSKGRSKENHNDGIVKSI  
 INGGAASKDGRIRVNDOLIAVGBSLGKTNDQDAETLRSMSTEGNKGMLQILVLR  
 RIISKNEKLSPPSPGPELPITATLDDERKISHLSYSGIEGLDESRRNALQILVLR  
 KIOLSPVVMKPPDDVITLEDRLPVLPPLHLSQSSSSSHDDVGVPTADACTWAKAIS  
 DSADSLSPDVPVLAFOREGFRQSMSEKRTKQFSDASQDLPVKTTRKSSMDLGLAD  
 ERLKNTVDQKAGSPSRDVGSLGLKSSLSLEISLOTAVAVETLNGDIFHRPRLIR  
 GRCNESEFPAALIDKSYDRPAVDDDEGETLEEDTESSSGRESVSTASDQSHLE  
 R  
 ROMNNOEGDKTDKDKTKGKEKKKDKDKMKKMKMLGLDQMSLAKIPKPK

BASE COUNT 1003 a 811 c 934 g 721 t 1 others  
 ORIGIN

Query Match 53.6%; Score 3050.6; DB 9; Length 3470;  
 Best Local Similarity 99.6%; Pred. No. 0;

	Matches	3072;	Conservative	0;	Mismatches	4;	Indels	9;	Gaps	1;
QY	1	atgaagaagtcg	atcgtctgcgaagcagaccgggtgcgtgcgtgcggtgcgaagcagccac	60						
Db	288	ATGAAAGTACCG	ATGCTTGGTGGAGGAGACCGGGTGGTGTCCGTTGGGGGAGCGGCCAC	347						
QY	61	atgaagcttc	atcgaagcctatccagcagcggtgacgcgctacccgaagagccatgcgcaag	120						
Db	348	ATGAAAGTTTTCAG	CTCATCCAGCAGCGCGGTGACCCCTACCCGGAAAGGCCATTCGCCAAG	407						
QY	121	gatccaac	atcctcgtatccaagtgatcgtcttggaaacatgagatggaggaataactaagac	180						
Db	408	GATCCAAACTCTAG	TATACAGGTGCATGGTGCATGGTAACATGGAGATGGAGGAATTAATAGAC	467						
QY	181	cttgatgc	atcctctctgtgcatgtaacagataaagaacagactgtagcaatgcttgat	240						
Db	468	CTTGATGACATTTCT	TGTGTGATGTACACAGATTAAGACAGCTGTGACAGTGTGTTGAT	527						
QY	241	gagcagag	atccacatccagcaggtgtagtcacagctgcgaatgccagttccacggtatccacagagc	300						
Db	528	GAGCGAGTATCCAC	ATCATCAGGAGTGTGGCACCAATGTCACAGTTCACAGGGTATCCAGAGAC	587						
QY	301	ccagagata	atcttggtagtgagcttggcaccacaacaaatgctcaagccttcagccttaccacaa	360						
Db	588	CCAGGAGATATTTGG	TATGGTAGCGCTTGGCACCAACAATGTCTCAGCCCTTCAGCGCTTACCAA	647						
QY	361	gcaaaag	tgaatatgtagtcaaccctcagctcccttcgagcaaatatgccttcattgt	420						
Db	648	GCAACAAATGAAAT	TGAGAGTCAACCTTCAGCTTCGAGCAAAATATGCTTCTTATGTT	707						
QY	421	cgaagcag	tagtgcacacagctctaattggtgcctcctcactctctgcatgtagtaattt	480						
Db	708	CGACCACTAATGAC	CCAGCTCTAATTGGCCCTCCACTTCGTCAAGTATAGTAATTTT	767						
QY	481	tcctctga	agagccttcaagaaataatccacaacgcctggtcaacaaagcctgctctc	540						
Db	768	TCTCTGTGAAGAC	CCCTTCAGGAAATATCCCAACCTCTGTCTCAACAACAGCTGGCTTCCTC	827						
QY	541	aagcgaac	acactgctcggagtccttaaaacccctgcgacgaagaagaatgaaaacctacaga	600						
Db	828	AAGCGAACA	CTGCTGGGAGTCTTAAACCTGCGACAGAGAAAGATGAAACTTACAGA	887						
QY	601	agcctccc	cggtatatactagtaactggtcttaacccaattcagagagacaatgctgcgtcg	660						
Db	888	AGCCCTCCGCGG	AGTAAGTAAGTCTTAACCAATTTACAGAGACAAATCTCGCTCG	947						
QY	661	tctctga	gtgcagctaccccaatggtggtgcgaagtgcctgcgagaaacaagaagatatgag	720						
Db	948	TCTCTGACTG	CCAGTACCCCAATGGTGGGCAAGTGGCGGAACAAACAGACGATGAG	1007						
QY	721	gatggaga	agaagagatataacagctgttgaacctgtttgacatgcctgcgaacaggtttg	780						
Db	1008	GATGGGACAGAA	AGAGATTAACAGTGTGTGAACCTGTTGGACATGCTGACACGGGTTTG	1067						
QY	781	gagcata	atacccaacttctctcgtatgataatgtaaaagctcgtlaagaatccccaagat	840						
Db	1068	GAGCATATACCA	CAATTTCTCTGATGATATGTAAGCTGTGAAGTCCCAACAGAT	1127						
QY	841	ggaggg	cctctcggaatccatgtagtgccttcagtgctcgagcggcgcaagaacctgag	900						
Db	1128	GGAGGCGCTTGG	GAATCATGTACTGTGCCCTTCAGTGTCTGAGCGGCGCAAGACCTGGGG	1187						
QY	901	ttattag	taaaaagattggagaaggtggtaagcttgacaatgaaatccttttcgttag	960						
Db	1188	TTATTAGTAAAN	CGATTTGGAGAAAGTGTAAAGCTGAACATGAAANAATCTTTTGTGTGAG	1247						
QY	961	aatgat	tgcatgtgaagataatgatgtgcgaaccttcgaaatagaagaattgaaacaaga	1020						
Db	1248	AATGATTCAT	GTGCAGATTTGATATGAGCTTCGAAATATGAAGATTTAACAACAGA	1307						
QY	1021	caacata	gtcttcgcaagcaatgcgtacacccaatcatcttggctcaatgtagtcttcgtca	1080						
Db	1308	CACATATATG	TTTGGCCAGGACCATGTGGTACACCATCATATTTGGTTCACATGTCTCTGCA	1367						

OY	1081	gcaataaagagcagctatgaaacaatactcccgaagtggagaagaacaatactattccaagc	1140
Db	1368	GCAAATTAAGAGCGATGTGAACAACATACTCCAAAGGAGAAACAAATTAATAATTCAGC	1427
OY	1141	cgcttcagccctgcacacagatataatgaaacaagagctgtaaacagctgcaggtctaac	1200
Db	1428	CGTTTACCCCTGACACCCAGCTATATTGAAACAGAGGTGTACACAGTGCAGGCTTAC	1487
OY	1201	acggltgcagagagcaccgccagctgaaaccaccgccttgacagatagactctcaactaaga	1260
Db	1488	ACGGTGCAGAGAGCACCCCGCACTGAACACACCCCGCTGACACATATAGACTCTACACA	1547
OY	1261	ctacctatagcgacacccctctggaaaaaacacacactctgcctccagagccctgcagctacg	1320
Db	1548	CTACCTCATATACGCCACACCCCTGGGAAAAACACATCCGCTCCAGCTCCGCGACCTTAG	1607
OY	1321	aatgtattagtaagactgtgaacgactgtgtatataacaccaaanaaaatagcaagaagctt	1380
Db	1608	AATGATTAATTACTAGACTGTAAAGAGTGGTTATTAACACCAAAAAAATATAGCAAGGCTT	1667
OY	1381	aatatccagcttaagaaagttacagaagtttggatltcaagcatcacttccagagatga	1440
Db	1668	AATATCCAGCTTAAAGAAAGGTATACAGAAAGTTTGGATTCAGATCATCTTCCAGAGATTA	1727
OY	1441	acaaatagttgcctcagctcccaatcattatgttaaaacattctccccggggggcgagcat	1500
Db	1728	ACAAATAGTGGCTCAGCTCCACTATCTATGTATAAAACAAATCTCCCGGGGGGCGGCA	1787
OY	1501	caggatgcgcgaacttaagcagagagacagactataaggtaaatlgagtagattagtg	1560
Db	1788	CAGGATGGCCCACTTAAGGCGACGAGACAGACTTATAGAGTAAATGAGATTAATTTAGTG	1847
OY	1561	ggcaatccccaagaggaattgtttctgcgtgtgaagaacacaaagtgaagaagaacttg	1620
Db	1848	GGCAATCCCAAGGGAAGTGTGCTTGCTGTGAGAAACCAAGATGGAAGAACTGTG	1907
OY	1621	agcctctcgtgctcttcgcagagaaagccttccaccaagaagaaactgtaatgcagagca	1680
Db	1908	AGCCTTCTGGCTTTCGCCCGAGAAAGAGCGCTTCACCCCAAGGAACTGAATGCAGAGCA	1967
OY	1681	agccaagatgcagatcccaaaaagaagaagcagaagatggagatattgtttctacact	1740
Db	1968	AGCCAGATGCAGATTTCAAAAAGAAACCAAAAGCAGAAAGATGAGATATTTGTTACACTT	2027
OY	1741	gatgcacccaaggaacttcgcacattggaatgcacactaaatgatcagactctgcagc	1800
Db	2028	GATGGCACACAGGAATTTCTGACATTTTGAAAGTCCCACTTAATGATTCACAGATCTGCAGC	2087
OY	1801	cttggtgtcagtgtcacaagttaacagtcgaagaacccaacgcagacttggaaactctt	1860
Db	2088	CTTGGTGTCACTGTCCAAAGGTAAACCGGTCCAAAAGAACACAGCAGATTTGGGAATCTTT	2144
OY	1861	gtcaagttccattatataatgagagcagacatcnaagatggaaaggtcttgggtgaatgat	1920
Db	2148	GTCAGTCCATTATTAATGCGAGCGACGACTCTTAAGATGGAAGCTTCGGGTGAATGAT	2207
OY	1921	caactgatcagtaaaatggaagaatccctgttgggcaagaacaaaccaagaatgcataaggaa	1980
Db	2208	CAACTGATAGACGATTAATGAGATCCCTGTTGGCAACGACAAACCAAGATGCCATGGAA	2267
OY	1981	accctaagaaggtctatgcttactcgaaggaaataaacggggaatgtatccagtatttgt	2040
Db	2268	ACCCTAAGAAAGGTTATGTCTACTGAAAGGCAATAAACGAGGATGTATCCAGTTATTTGTT	2327
OY	2041	gcaagggaaataagcaagctgcaatgagctgaatgcacccggagccccctgcagactgag	2100
Db	2328	GCAAGGAAATTAACCAAGTGCATGAGCTGAATCACTACCTGGGAGCCCCCTTGACACTGAG	2387
OY	2101	ctgcccatgaaacagcgtttgatatagagaagaacgaagaattcccatctcccttaagct	2160
Db	2388	CTGCCCATTTGAACAGCTTGATGATAGAGAAACGAAGATTTTCCATCTCCCTCTACACGT	2447

QY 2161 gggattgaggggcttgatgtaacgcccagcaagaatgctgcctccagtagaataatggt 2220  
 Db 2448 GGGATTGAGGGGCTTGATGTAATCCCCAGCAGAAATGCTGCTCCTAGAGTAATG --- 2504  
 QY 2221 gagtcagtaataaccagctgtccctacagtaataatgcccagaatgtaacatgtaac 2280  
 Db 2505 -----GTTAAATTCACAGCTGTGCTCCCTACAGTAAATGCCCCAAGTATGACATGTCAT 2558  
 QY 2281 atagaagaatgacaggtgtgacagtgcttccctcacatctctctgacagctcctccag 2240  
 Db 2559 ATAGAAGATGACAGCTGTGCTCCAGTCTCCACATCTCTGACACAGCTCTCTCCAGC 2618  
 QY 2341 tcccatatgatgtgaggtgtgtgagggcgagatgctgtgtaacttggcccaaggtcgaac 2400  
 Db 2619 TCCCATATGATGTGAGGTTGTGAGGGCAGATGCTGTGTACTTGGGCCAAGGCTTCCAGC 2678  
 QY 2401 agtgatcagcagctgtcttctgagtcagatggttgaatccagttcttcttcaaga 2460  
 Db 2679 AGTGATTCAGCCAGCTGCTTGTGAGTCCAGATGTCAGATTTCTTCTTCAACGA 2738  
 QY 2461 gaagagttgagctgacaggtatgtcagaanaaagcacaagaatcttcagatgccagt 2520  
 Db 2739 GAAGGATTTGGACGTCAGATGTCAGAAAACGACAAACAAATTTTCAGATGCCAGT 2798  
 QY 2521 caatlgatctgtaaaacacagaatacaaaaagcatgagtttggatgtaagtcagag 2580  
 Db 2799 CAATGATTTGCTTAAACACAGAAATCAAAAAGCATGATTTAGGTATGATGCTGACAG 2858  
 QY 2581 actaaactaatcagtgatgtaacagaagaagagttctccacagagatgtggtcct 2640  
 Db 2859 ACTAAACTAATAACATGATGATGACAGAAAGAGGTTCTCCACAGAGATGTGGTCT 2918  
 QY 2641 tccctgagctcgaagaagtcagagctcgttggagagctgcagacagctgtgcagagtg 2700  
 Db 2919 TCCCTGAGCTCAGAAAGTCAGCTCATTTGGAGTCTGACAGCCGAGTTGCCAGGCTG 2978  
 QY 2701 acttggatgggagatattcccttccatcgtccacggcgagatatacagagggagga 2760  
 Db 2979 ACTTTGAATGGGATATTCCTTCCATGTCGACGCGCGCGATATATAGAGGAGGGA 3038  
 QY 2761 tgcattgagagcttcaagctgtgcacatgcaaatcttataaaccgcggtatgat 2820  
 Db 3039 TGCAATGAGAGCTTCAGAGCTGCCATTCGACAAATCTTATGATTAACCCGCGGTATGAT 3098  
 QY 2821 gatgatgaagcatggaagcttgaagaagaacacagaagaagttaagaatcagggaga 2880  
 Db 3099 GATGATGAAGCATGGAACCTTGGAGAAACACACAGAAAGTTCAAGTCCAGGAGA 3158  
 QY 2881 gagctgtatccacagcagtgatcagcttcccaatctctgagagacaaatgaatgga 2940  
 Db 3159 GAGCTGTATCCACAGCCAGTGAACCTTCCCACTCTGAGAGACAAATGAATGCA 3218  
 QY 2941 aaccaagaagaagtgatgaagctgtatagaanaaagataaactgtgaagaagaagaag 3000  
 Db 3219 AACCAAGAAAGTGATTAAGACTGATAGAAATAAGATTAATCTGAAATAAGAAAGAG 3278  
 QY 3001 aaagatgagataagaagaagataaatagaagccagaagaagggaatgctgaagggttg 3060  
 Db 3279 AAAATGAGAGATAAGAGAGGATAAATGAAGCCAAAGAGGAATGCTGAAGGCTTG 3338  
 QY 3061 ggaagacatgttcaggttggcacaac 3085  
 Db 3339 GGAGACATGTTTCAGCCCTTGCCCAAC 3363

RESULT 11  
AY026057

LOCUS AY026057 4167 bp mRNA linear ROD 20-FEB-2001  
 DEFINITION Mus musculus PAR-3 180 kDa isoform (Par3) mRNA, complete cds.  
 ACCESSION AY026057  
 VERSION AY026057.1 GI:12965348  
 KEYWORDS house mouse.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 4167)  
 AUTHORS Lin,D., Gish,G.D., Songyang,Z. and Pawson,T.  
 TITLE The carboxyl terminus of B class ephrins constitutes a PDZ domain binding motif  
 JOURNAL J. Biol. Chem. 274 (6), 3726-3733 (1999)  
 MEDLINE 99121117  
 PUBMED 9920925  
 REFERENCE 2 (bases 1 to 4167)  
 AUTHORS Lin,D., Edwards,A.S., Fawcett,J.P., Moamali,G., Scott,J.D. and Pawson,T.  
 TITLE A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and apkc signalling and cell polarity  
 JOURNAL Nat. Cell Biol. 2 (8), 540-547 (2000)  
 MEDLINE 20394297  
 PUBMED 10934475  
 REFERENCE 3 (bases 1 to 4167)  
 AUTHORS Lin,D.C. Jr. and Pawson,T.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JAN-2001) Programme in Molecular Biology, Samuel Lunenfeld Research Institute, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

#### FEATURES

source location/Qualifiers

#### gene

1. 4167  
 /organism="Mus musculus"  
 /strain="N1H Swiss"  
 /db\_xref="taxon:10090"  
 /gene="Par3"  
 /note="Par3; scaffolding protein"  
 /product="PAR-3 180 kDa isoform"  
 /protein\_id="AAK07669.1"  
 /db\_xref="GI:12965349"  
 /translation="MKVYCFGRPPVPCDGRMKYFSLIQAVTRYKAVANDPNY

#### CDS

/codon\_start=1  
 /product="PAR-3 180 kDa isoform"  
 /protein\_id="AAK07669.1"  
 /db\_xref="GI:12965349"  
 /translation="MKVYCFGRPPVPCDGRMKYFSLIQAVTRYKAVANDPNY

BASE COUNT 1145 a 1079 c 1174 g 769 t  
 ORIGIN

Query Match. 51.9%; Score 2952.2; DB 10; Length 4167;  
 Best Local Similarity 83.4%; Pred. No. 0;  
 Matches 3442; Conservative 0; Mismatches 613; Indels 72; Gaps 5;

QY 1 atgaagtgacgctgtgtcttcgagcagccgggtgtgtgtcgtgagcagccac 60  
 Db 110 ATGAAGTGAACCGTGTCTTCGGAGGAGCCCGGCTGTCTGCGGAGATGCGCCG 169

QY 61 atgaagtttcagcctcatccagcagcggtgacccgctacccggaagccatcgcgaag 120  
DB 170 ATGAAAGTTTCAGCCTCATCCAGCAGCGGGTACCCCGTACCAGGAGCCGTGGCCAA 229  
QY 121 gatccaactactgtatcaggtgtcatcgcttggaacatgagatgaggaatactagac 180  
DB 230 GATCCAACTACTGAGTACAGGTGATCGCTGGAGCATGGAGATGGAGGATTTCTAGAC 289  
QY 181 ctgtatgcatctcttctgtatgtatgacagatgaagaacagctgtagcaatgttcat 240  
DB 290 CTGATGATCCTCTGTGTGACCTTTCATGACAAAGACAGCTGTAGCACTATTGTAT 349  
QY 241 gagcagatccacatcagcagatgagatgagcagctgacagctcagcaggtacagagc 300  
DB 350 GAACAGATCCCAACCATGAGAGATGATACACAGCGCAGCTTACAGGGAACCCAGAT 409  
QY 301 ccagagatattgtgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtat 360  
DB 410 CCAGAGATATTCGAGTACAGGTGGGACCAACATGTTCTGCTTTCAGCCTTATCAA 469  
QY 361 gcaacaagtgaattgagtcacacccctcagctcctcagcagcaatattgccttcatt 420  
DB 470 GCCAATATGAATTTGAGGTACCGCTTTCAGTCTTCGGGCAATATGCTTTCATGTC 529  
QY 421 cgaagcagtagtgcac 480  
DB 530 CGCGGACAGAGACCCAGCTTAACTGAGCTTTCACCTTGTGATGATTAACAATT 589  
QY 481 tcccttgagagccttcaagagaataatccacacacacacacacacacacacacacac 540  
DB 590 TCCTTACAGAGCCCTCCAGGAAAAACCCAGCCCTGTCACACACACCTGGCTTTC 649  
QY 541 aagcagacactgctgaggtctcctaaccctgcagcagagaagaagaatgaatacaga 600  
DB 650 AAGCAACACCCGCTGGAATCCCAAACTCCGACAGAGAAAGATGAAAACTTACAGA 709  
QY 601 agcctccgaggatagtaactgtgtcctaaccattcagaagacaaatgctgcgtcgc 660  
DB 710 AGCCTTCCACGGGATCCCAAGTGTGTCACACAGTTCACAGCAGACACCCGCTCC 769  
QY 661 tctcgtgagtcagtcac 720  
DB 770 TCCTGAGACGGCAGCCCAATGATGATGACCGGTGTGAGAACCAACAGAGATGAG 829  
QY 721 gatggagacagaagaatgaatacagctgtgtgaacatgtgtgaacatgtgtgaacatgt 780  
DB 830 GAAGCACAGAGAGAGACAGCAGCGAGTGGAGCGGTGGACATGCTGATACCGATTC 889  
QY 781 gagcatatacccaacttctcctcctgaatgtatgtatgaagctcgtgaagctcccaac 840  
DB 890 GAGAACATGCCCCAATCTTCCCTCATGATGATGATGATGATGATGATGATGATGAT 949  
QY 841 ggaagagcctcgtggaatcatatgtatgtatgtatgtatgtatgtatgtatgtatgt 900  
DB 950 GGAGGGCCCCGTGGAGATCCATGATGATGATGATGATGATGATGATGATGATGATG 1009  
QY 901 ttattatgaataacagatgagagaaggtgtaagctggaacatgaaatcttctcgtgag 960  
DB 1010 TTGCTTAGAGAGCGGTGGAGAAAGCGGTAGGCTGACCAAGAAACCTTTTCCATGAG 1069  
QY 961 aatgatgtcatgtcagagatgaatgagcagcctcgaatagagaatgtgaacagca 1020  
DB 1070 AATGATCTCATGTGAGATTTAAAGATGAGATCTTCGAAACAGAAAGATTTAGCA 1129  
QY 1021 caaatatggttcggcaagcagctgtgacacccatcatttgttcatgtgttctgtga 1080  
DB 1130 CAACATATGTTCCGCAAGCTATGCGTGGCTGCTCATTTGTTGTTGTTGTTGTTGTT 1189  
QY 1081 gcaataaagagcaglatgaacaactatcccaagatgagaagaacaatcattcatcaac 1140  
DB 1190 GCAAAACAAGAGACATATGAACTGTGTCACACCGAGAAACAAACACTACTCTCCAG 1249  
QY 1141 cgtttagccctgcagcagcagatattgaacaagagatgtgacagctgcaggtctcac 1200  
DB 1250 CGCTTACAGCCCTGACAGCCACTGTGTGCCAACAGAGATGTGGCAACATGCCCCCT 1309  
QY 1201 acggtgcagagagcaccctgacttgaaaccccgctgcagcagatagaaactcctaaga 1260  
DB 1310 GCATTGCCAGAGACACCCAGACTGATGACGCCAGACAGCTGATGATCTCACCCCA 1369  
QY 1261 ctacatcagcagcaccctcgggaacacacacacacacacacacacacacacacacac 1320  
DB 1370 CTACCTCATAGTGTCTACGCTTCAACCAACACCCGAGCCCGGCTTGTGCTCACCC 1429  
QY 1321 aatgtatttagtgcagctgtaagcagtggttatacaccaaaaaatagcagaagctt 1380  
DB 1430 AGTGTCTTAGTACCAACGTAGACAGTGTACCAACAGCAAGAAATAGGACAGAGCT 1489  
QY 1381 aatattccatgaagaaggtacagaaggttggagttcagatcactccagaatgtga 1440  
DB 1490 AACATCCAGCTTAAAGAGTACAGAAAGCTGGGATTCAGCATCTCCGGAGCTG 1549  
QY 1441 acaatagtggtcagctcacaatctatgtgaataacatctcccccggggggcgcatt 1500  
DB 1550 ACCATAGTGTGCTCAGCTCCCATTTATGTCAGAAATATCTTCTGTGAGGGCTGCCAT 1609  
QY 1501 cagatgcgcagacttaagcagagacagactatagagtaaatgagtagatgtatgt 1560  
DB 1610 CAGATGCGAGACTTAAGCAGAGACCGCTAATAGAGTCAATGAGATGATTTAGCA 1669  
QY 1561 ggcgaatcccaagagaatgttctcgtgtgtgagaagcacaagatgagaagaaactgt 1620  
DB 1670 GGCAAATCCAGAGAGAAATGTTTCCCTGTGAGAGAACCAAGATGGAGGAGCTGT 1729  
QY 1621 agccttctgtcttctgcagagagacacacacacacacacacacacacacacacac 1680  
DB 1730 AGCTTCTGTGCTTGTGTGACAGAGAGCTTTCACCAAGGAATGATGCTGAACCA 1789  
QY 1681 agccagatgcagatctccaaagaacagaaagcagaatgagatgtgttcttaacct 1740  
DB 1790 AGCCATATCAGACTCAAAAGAAAGAAAGCTGAGATGAGAGAGCTGTCTCACACC 1849  
QY 1741 gatggacacaggaattctgacattgaaatgtcccaattatctcaagatctgaagc 1800  
DB 1850 GATGATACAGGAGATTTCTGACTTTGAAATTCCTGATATACACAGATTCGAGAG 1909  
QY 1801 ctgtgtcagtgatcaagaagtaacacacacacacacacacacacacacacacacac 1860  
DB 1910 CTGTGTGATGATCAAGGGAACGTTCCAAAGAAACACAGCATTTGGGAGTCTTC 1969  
QY 1861 gtaagtcattatgaatgagagagcagcatcctaagaatgagagcttcgggtgaatgat 1920  
DB 1970 GTTAAATCATTTATCAATGATGAGTGCATCTTAAAGATGAGAGCTGAGGTAAATG 2029  
QY 1921 caactatagcagtaaatgagaatccctgttgaggaagaacaaacgaatgcatggaa 1980  
DB 2030 CAGCTGATAGCTGTGAATGAGAGATCTTACTGAGCAACCAACAGAGGACATGAG 2089  
QY 1981 accctaaagagctctatctgaagcagaaataacacagagaatgtatccagcttatgt 2040  
DB 2090 ACTCTAGGAGAGCATGTGTCACAGGAGGCAACAGCTGTGATTCAGATCATTTG 2149  
QY 2041 gcaaggaagaaagcagatgcaatgagcagcagcagcagcagcagcagcagcagc 2100  
DB 2150 GCGAGCGGATCGAGATGTAAGAGAGCTGCGGTCTCTTGGAGCCCTGTGCACTGAG 2209  
QY 2101 ctgcccattgaacacagcgttggatgataagacaggaattcccatccctcagat 2160  
DB 2210 CTGCCCATGAGACAGAACTGAGTACCGAGAACGAGATTCACACTCCCTCTACAT 2269  
QY 2161 ggaattgagagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2220  
DB 2270 GGGATGAGAGGGGTGATGTAGTGTGCGCCACAGAGACAGCACTGAGAGGATATG -- 2327  
QY 2221 gagtcaagtaataaccagctgtccctcaagatgataatgcccagaatgacactgtcatt 2280

```

Db 2328 -----GTAAATGCCAGCTGTCCACGTCACATGCTCAGATGACACTGTCAG 2380
QY 2281 atagaagatcacaggttgcagctgtctccacatctctccacagctctccacag 2340
Db 2381 ATTGAAGATGACAGGCTGCTGCTGCTCCTCCTCCTCCTGACAGTCTCCTCCTC 2440
QY 2341 tcccatgatgatgtgggttctgtacagcagatgctgactgtggtgccaagctgcaac 2400
Db 2441 TCCCATGATGATGCTGATTCATATGACAGAGGACGACGCGGCCAAGATGACATC 2500
QY 2401 atgatcagcagcagctgctcttgatccagatgttgcagcttctgtcttcaaga 2460
Db 2501 AGGACTCAGCGACTGCTCATTTGATGACAGTGTGATCCGCTTCTCCTTTCAACG 2560
QY 2461 gaagagattgagacgtcagatgctcagaagaaacagcaagaacatttccatccagt 2520
Db 2561 GAAGGATTTGAGAGCGCAGATGTCAGAAAAACCAACCAATTTTCAAGATGCCAGT 2620
QY 2521 caattgatttgcgttaaaacacagaaatcaaaagcatgatttagtataagctgacag 2580
Db 2621 CAATTGGATTTGCTTAAAAACAGAAAAATCAAAAAGCATGATTTAG---TACTGACG 2677
QY 2581 actaaactcaatcacagttgatacagaagaagaggttctccagacagagatgtggtcct 2640
Db 2678 ACTAAACTCAATCACAGTGTGATACAGAGAGCAGCTCTCCAGTAGAGATGTGGAGCC 2737
QY 2641 tccctggtctcgaagaagtcgaagctcgttctgagagctcgcagacagctgtgcagtg 2700
Db 2738 TCCCTGAGCTCTGAAAGAAATCTGCTCTTACAAAGTCTGACAGACCGCTGCGCCAGTG 2797
QY 2701 acttgaatgagatctctctctctctctctctctctctctctctctctctctctct 2760
Db 2798 ACGCTGAACGGGAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2857
QY 2761 tgcagtgagcttcagagctgcgcacacatctctctctctctctctctctctctctct 2820
Db 2858 TGCAACGGAAGCTTCAGAGCGCGCATGTGACAAAGTCTGACAAAGCCATGCTTGATAT 2917
QY 2821 gatatagaagcagaggaaccttggaaagaacacagaagaagtlcaagatcagggaga 2880
Db 2918 GACGACGAGGATGAGACACTTGTGAAAGACACAGAGGAAACCTGACGCTCAGGAGG 2977
QY 2881 gactctgtatcacacagctgcagctctccacatctctctctctctctctctctctctct 2940
Db 2978 GAGTCCGCTGCTCAGCTCAGGATCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3037
QY 2941 aaccaagaagaagtgataagactgatagaagaaagaaagaaagaaagaaagaaag 3000
Db 3038 GACCCAGAGAAAGGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3097
QY 3001 aaagatagataagaagaagataaatagaagccagaagaagaaatgctgaaggcttg 3060
Db 3098 AAAGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3157
QY 3061 ggaagacatgttcagggttctgcaaacatcgaagaaagatacaagattgagaagaaagcggat 3120
Db 3158 GGGGACATGTTCAAGGTTTGGCAAAACATGCAAAAGATACAAAGATGAAAAAATGGGTGCA 3217
QY 3121 ataaaaatacaagatctcttaatacaagaagaagagagatacgaatgaagcagagcag 3180
Db 3218 ATAAAAATTCAGAGATTTTCACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3277
QY 3181 gagaagattcaagccaaactcgaagatttaggaagcagaagcagaagcgtgactat 3240
Db 3278 GAGAGGATTTCAAGCCAAATCTCAGAGATTAGGAGAGCGGCAACCGGAGCGCTGACTAT 3337
QY 3241 gctgaatcaatcaagatttcaatcagacattgtgctgagatgagatgaatgataagagga 3300
Db 3338 GCAGAGATTCAGAGATTTCACTGAGAGGTTTGGCTGTGATGACAGAGTTGGCTATGGGGGC 3397
QY 3301 gtctctctataaaggttccatgctctcaagctcagacctcagagcccaagagaagg 3360

```

```

Db 3398 ATGTATCTTATGAAGGCTGCTTGGCTCTCAATGCCAGACCCAGAGCCCAAGAGAGGG 3457
QY 3361 catatgatgatgactgttctatagcccaagtcgaagacgcgagatctccaacccctcact 3420
Db 3458 CACCTGATGAGACACTTGTGATGCTCCAGATGAAGAAACCTCGAGAGTCCAA-----ACCT 3511
QY 3421 gtagcagtaaaagttcaactctctcagaataatgataagctgataagcgtctgaggaaga 3480
Db 3512 GGAGACAGCAATCTGACCTCTTACCAACCATGACCGGATACAGCGTCTACGGCAAGAG 3571
QY 3481 ttccagcaagaacagaatgaagaatgataagaatgctggcgagactatagttttgag 3540
Db 3572 TTCAGCAAGCAACCAAGATGAGATGTGCAAGACCGGCGCCGACTCTACAGCTTTGAG 3631
QY 3541 caacctggccggaacgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3600
Db 3632 CAATCTTGTGCTCAGCTCTCGGCGCGGCATGCGAGAGTGTGCGGACTGCTCGGTGGAG 3691
QY 3601 gtgcagatgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3660
Db 3692 GTTCAAGTACAAACGCGAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3751
QY 3661 tacagctctcgtcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3720
Db 3752 TACAGCTCAGTCCCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3811
QY 3721 gagcagaactctctcccttgaggaaagcttccagatgccaagaagaaacccagttactc 3780
Db 3812 GAACAGAACTACGCGCTCTGAGAGGCTTCCAGAGTCCCAAGAGAGAGAGAGAGAGAGAGAG 3871
QY 3781 agtaccagagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3840
Db 3872 AGTACCAAGGCTCAAGAAATGCTATCTAGCGGAGACGCTTTAATGCGAGAGGCTATG 3931
QY 3841 ctggaactcaagagctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3900
Db 3932 CTGAGAGCCAGAGGCTTCTCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3991
QY 3901 caagctctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3960
Db 3992 CAGCTCTCAGCTATGAGATGACAG----- 4015
QY 3961 taagccctcccaagggccctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4020
Db 4016 -----GGGCCCTTCGGCAAGATGTGCCCTTCTCTCATCTCAGGTGGCT 4060
QY 4021 aggttaaaagacttaagactcctgagaagaagggccctctatctctgagcagcaaa 4080
Db 4061 AGGCTAAACAGACTGTGACACCGGAGAAAGGCGCCCTTCACTCTGAGCTCAAGA 4120
QY 4081 taacgg---atgctcatgtcgcgcgaataaagaacatttctctatga 4124
Db 4121 AGGATGGCAGCATTGTGTGATGCAATTAAGGACATTTATCCCATGA 4167

RESULT 12
AF332592 2967 bp mRNA linear PRI 16-NOV-2001
LOCUS AF332592
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein
ACCESSION AF332592
VERSION AF332592.1 GI:14579308
KEYWORDS Down-regulated expression of atypical PKC-binding domain deleted
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2967)
AUTHORS Fang,C.M. and Xu,Y.H.
TITLE Down-regulated expression of atypical PKC-binding domain deleted
JOURNAL Cell Res. 11 (3), 223-229 (2001)
MEDLINE 21520080

```

PUBMED 11642408  
REFERENCE 2 (bases 1 to 2967)  
AUTHORS Fang, C. and Xu, Y.  
TITLE Exon/Intron Structure and Splicing Variants of A Novel Human Polarity Gene, HASIP  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 2967)  
AUTHORS Fang, C. and Xu, Y.  
TITLE Direct Submission  
JOURNAL Submitted (29-Dec-2000) Laboratory of Molecular and Cellular Oncology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, P. R. China

## FEATURES

source  
1..2967  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/map="10p11.2"  
1..2967  
/note="ASIPSB; alternatively spliced; similar to GenBank Accession Number AF196186"  
/product="atypical PKC isootype-specific interacting protein short variant b"  
/protein\_id="AAK69192.1"  
/db\_xref="GI:14579309"  
/translation="MKTVVCFRTVVVPCGDGHHKVPFLIQAVTRYRKAIKADPNY  
MIOVRLHKGSGTLLDLDLDCVADDDKDRVAVRBDPDDHGGDGTASSGTCQSP  
TEGSELGNNVSAFOYATSEIETVPLVLANPDLHRRSDPALTLGISTVSQNF  
SSEPSKRNPTKMSSTAGFLKONTAGSKTCDKDKEDVRSLLPRTSMWSNOFQDNA  
RSSLSASHPWKNLKEQEDQEDTEEDNSVEPEVGHADTGLEHLPNSLDMVLE  
VPNGGPGLGIVHPFSPARGKTLGLVRLKRGAEHNLREMDCLVRIINDCLRN  
RRFOAHMFROAMRPTIIPFVVPAAKEDYEOLESEKNVYSRSPSOYIDNR  
SVNSAGLHTVORARLNHPROIDSHSLPSAHPSGKPPAPAPAPQNVSTVSSG  
YNTKKIKRLNIQAKKEGFEFTSDVTIGSAPITYVKNILPRGAIIDGRKAK  
DRLLEKNGVDLVKSQSEVSVSLRSTKEGVSLVFEQEDAFHRELKAEDEDTVLT  
PDGTREFTEFPLNDSGAGLVSVKRNREKNADIGIVKSLINGANSKDLRL  
VNDLIANGNSLIGKTNDAMETLRSMSTEGNKAGIQLIVARIKCNLKPQS  
PGPELPLETALDRERISHSIVSGSHDVGVTAGVWAKAAISDSADSLSPVOD  
DTVLIEDRLPLVPLHSDSSSHDVGVTAGVWAKAAISDSADSLSPVOD  
VIAFORBEGQIADERTKLVYDDOKASPSRDVPSGLKSSLESLQTVAEVTL  
NGDIPFHRPRITIRGCSNESFRALIDKSTDKPAVDDDEMETLEEDTESSSGR  
ESVSTASDQPSHSLERDMNGNOEKDKTKDKTKGKKRDKRDKEMKAKKMKMLK  
GLDMFSLAKLKPERK"

CDS  
BASE COUNT 830 a 653 c 776 g 648 t  
ORIGIN

Query Match 48.7%; Score 2772.6; DB 9; Length 2967;  
Best Local Similarity 95.4%; Pred. No. 0;  
Matches 2943; Conservative 0; Mismatches 4; Indels 138; Gaps 3;

QY 1 atgaagagacggtgtgcttcggaagacccgggtggtcgtgcgtgaggaagcagc 60  
DB 1 ATGAAAGGACCGTGTCTCGAGGAGCCCGGGTGTCTCGCGGAGACGGCCAC 60  
QY 61 atgaagtttcagcctcatcagcaggcgggtgacccgctaccggaagcattcgccaa 120  
DB 61 ATGAAAGTTTACGCTCATCCAGCAGCGGTGACCCGCTACCGGAGGCCATCGCCAA 120  
QY 121 gatccaactactgatacaggtgcatcgcttggaacatgagatgagaaactaagac 180  
DB 121 GATCCAACACTGCGATACAGGTGATCGCTTGGAACATGGAGAGAGAACTACTAGAC 180  
QY 181 cttagtgacattcttcttgatgtagcagataaagacagcctgtagcagttgtat 240  
DB 181 CTTGATGACATTCTTGTGATGTACAGACGATTAAGACAGCTGTACCACTGTTTGT 240  
QY 241 gagcagatccacatcaggaaggtgtagcaccagtgccagttccacaggtaccagac 300  
DB 241 GAGCAGATCCACATCAGCGAGAGTATGGCACCACTGCCAGTTCCAGCGGTACCCAGAC 300

QY 301 ccagagatatltgtagtagcgttggcaaccaaatgtctcagccttcagccttaccaa 360  
DB 301 CCAGAGATATTTGTAGTAGAGCTTGGCACCAAAATGTCTACGCTTTCAGCCTTACCA 360  
QY 361 gcaacaagtgaattttagtcaaccttcagcttccttcaggaataatgtccttcatt 420  
DB 361 GCAACAAGTGAATTGATGAGGACACCTTCAGTCTTCGACCAAAATGTCTTCATGTT 420  
QY 421 cgaagcagtagtagcaccagctcaattgagccttcacactctcagtgtagtaattt 480  
DB 421 CGAGCAGTAGTAGCACCAGCTTAATTGGCTTCCTCCACTTCTCAGTAGTATTTT 480  
QY 481 tcctctgaagagccttcaggaanaaaatccacacagcttggtcaacaacagctgtc 540  
DB 481 TCCTCTGAAGAGCCTTCAAGAGAAAATCCACACGCTGGTCAACAACAGCTGGCTTC 540  
QY 541 aagcagaacactgtctggaggtctctaaacctgtgacaggaagaagatgaaactaca 600  
DB 541 AACGAGAACACTGTCTGGAGTCTTAAACCTGCGACAGAAAGATGAAAACCTACAGA 600  
QY 601 agctcccgaggatactagtaactgtgttaaccaatttcagagagacaatgtctcgt 660  
DB 601 AGCTCCCGAGGATACTAGTAACCTGTCTTAACCAATTTAGAGACAAATGCTCGCTCG 660  
QY 661 tctctgagtgccagtcacccaatggttgggaagtggtcggagaacaagaacagatgag 720  
DB 661 TCTCTGAGTGCCAGTCACCAATGTGGGCAAGTGGCTGAGAAACAAGAGATGAG 720  
QY 721 gatgggaacgaagaagataacagtcgtgtgaacctgttgacaatgctgacagggatt 780  
DB 721 GATGGGACGAGAAAGATACAGTCTGTTGAACTGTTGGCAATGCTGACAGGGGTTG 780  
QY 781 gaacataaccacacttctcgtgataatgtaagaactgtagaagccccaacgat 840  
DB 781 GACATATACCAACTTTCTCTCGATGATATGTAAAGCTCTGTAAGTCCCAACGAT 840  
QY 841 ggaaggcctctgggaatccatgtagtgcttcagtgctcgaagcgagacacccctgg 900  
DB 841 GGAGGCGCTCTGGGAATCCATGATGAGCCCTTTCAGTCTCGAGCGGACAGACCTGG 900  
QY 901 tctttagtaaaagatttggagaaggtgtgaaagctggaataaatactcttcgtgag 960  
DB 901 TCTTGTAAAGGATTGGAGAAAGGTGTAAAGCTGMAATATAATCTTTTCGAG 960  
QY 961 aatgattgcatctgcaagatgaatgtagtgagcacttcgaataagaatttgaacaaga 1020  
DB 961 AATGATTGCACTTCTCAGGATTAATGATGCGGACCTTGAATGAAATTTGAACAAGCA 1020  
QY 1021 caacataatgttcgcaagccatgctgtaacccaatcattgttccatgtgttcctgca 1080  
DB 1021 CAACATATGTTTCCGAAGCCATGCGTACACCCATCATTTGTTCCATGTGTCTCGCA 1080  
QY 1081 gcaataaagagagatgataaacaactatcccaagtgagaagaacaaatctattcaagc 1140  
DB 1081 GCAAAATAAGAGAGATGATGAACAATATCCCAAGTGAAGAACATTAATCTATTCAGC 1140  
QY 1141 cgttttagcccttcagcagcagatatttgacaacaggagtgtgaacagttgagaggttcac 1200  
DB 1141 CGTTTACCCCTACAGCCAGTATATTGACACAGAGTGTGAACAGTGCAGGGCTTCAC 1200  
QY 1201 acggtgcaagagacacccgactgaacaccccgctgtagcagataagactcactcaaga 1260  
DB 1201 ACGGTGACAGAGACACCCGACTGAACACCCCGCTGAGAGATATACACTCTCAAGA 1260  
QY 1261 ctaactcataagcagacacccctcggtgaacacacacatccgcttcagctcggacactcag 1320  
DB 1261 CTAACTCATAGGCGACACCCCTCGGAAAAACCACTTCGCTCCACCTCGGACCTCAG 1320  
QY 1321 aatgatttagtaacagctgtaagcagttgtttaacaacaaaataatggaagaggtc 1380  
DB 1321 AATGATTATTAAGTACGACTGTAAAGCACTGTATTAACCAAAAATAATAGCAAGAGCT 1380  
QY 1381 aatatacagcttaagaaggtttaggaattcagcatcacttcagagaagtga 1440



Db 1381 AATATCCACCTTAAGAAAGGTACAGAGTTGGATTCCACATCCTCCAGAGATGA 1440  
 Qy 1441 acataagtgctgcaccccaatctatgtgaanaacatctccccggggggccatt 1500  
 Db 1441 ACATAATGGTGGCTACGCTCCACATCTATGTGAAAAACATTTCCCGGGGGGGCCATT 1500  
 Qy 1501 caggatgacgaacttaagcagagacagacttataaggtlaattgagttagttatg 1560  
 Db 1501 CAGGATGSCCCTTAAGGACGAGACGACTTATAGAGTAAATGAGTAGATTTAGTG 1560  
 Qy 1561 ggcacatcccaagaggaagtgtgttcgtgttgaagacccaagaatggaagaaactgtg 1620  
 Db 1561 GGCATAATCCCAAGAGAGAGTTGTGTGCTGTGAGAAAGCACAAAGATGAGAAAGACTGTG 1620  
 Qy 1621 agccttcgtccttcgcagagaaagccttcacccaagagaaactgaatgcagagca 1680  
 Db 1621 AGCCTTCGTCTTTCCCGAGAAAGACGCTTCACCCAAAGGAACT----- 1667  
 Qy 1681 agccagatgcagatcccaagaagaagacagaaatgagatattgttcttaccct 1740  
 Db 1681 -----GAAAGCAGAAAGATGAGATATTGTTTACACT 1701  
 Qy 1741 gatggacagaggaattcttgcacatttgaagtcacacttaattgagttgcagagc 1800  
 Db 1702 GATGGACAGAGGAATTTCTGACATTTGAAAGTCCCACTTATGATTCAGGATCTGACAGC 1761  
 Qy 1801 ctgtgtgtcagtgcaaaagtaaacggtcacaagaagaaacagagatttggagatctt 1860  
 Db 1762 CTGTGTGTCAGTCAAAAGTAAACGGGTCAAAAGAAAGAAACGAGATTTGGGAATCTTT 1821  
 Qy 1861 gtcaagtcattatattatgagagagcagacatctaaagatggaaggttcggagatg 1920  
 Db 1822 GTCAAGTCCATATTAAAGAGAGAGACATCTAAAGATGGAAGCTTCGGGTGAATGTAT 1881  
 Qy 1921 caactgataagtaaatgagagaaatccctgttggcagaacaaacaaagatgcatgaga 1980  
 Db 1882 CAACGTAGACAGTAATAATGAGGATCCCTGTGGGCAAGCAAAACCAATGCGCATGGA 1941  
 Qy 1981 accctaagaaggtctatgtctactgtaagagcaataaaacggaatgagatttctt 2040  
 Db 1942 ACCCTAAGAAAGTCTATGTCTACTGAAAGCAATTAACGAGATGATCCAGCTTATGT 2001  
 Qy 2041 gcaagagaataagcaagtcagatgagtcgaagtcacccggagagcccccctggacatg 2100  
 Db 2002 GCAAGGAGATTAAGCAAGTGCATGACCTGACGTGAGGAGCCCTCGACCTGAG 2061  
 Qy 2101 ctggccattgaacacagcgttgcagatgataagagaacgaagaattcccatctccatcag 2160  
 Db 2062 CTGCCCATTTGAACAGCGTTGGATGATAGAGAAACGAAGATTTCCCATCTTACAGT 2121  
 Qy 2161 gggatggagggcgttgaatgaatcccgccagcagaatgtctgcccagtaagataatggt 2220  
 Db 2122 GGGATTTGAGGGGCTTGATGATCGCCGACGAAATGCTGCCCTCACTAGATGATG--- 2178  
 Qy 2221 gagtcaggttaataccacgttccctcagtaagtaataagcccaagaatgacacgtcat 2280  
 Db 2179 -----GGTAAATACCACTGCTCCCTACACTGAATATGCCCAAGATGACATGTCAAT 2232  
 Qy 2281 atagaagaatgacaggttgcagtgcttccctccacatctctgcacagtccttccacg 2340  
 Db 2232 ATAGAAGATGACAGGTTGCCAGTCTCTCCACATCTCTCTGACCAAGTCCCTTCCAGC 2292  
 Qy 2341 tcccatgtatgtgggggttgtgaagcagagatgctgtactgtggcgaagctgcaatc 2400  
 Db 2293 TCCCATGATGATGTGGGGTGTGACGGGAGATGTGTGACTTGGGCAAGGCTGCAATC 2352  
 Qy 2401 agtgcattcagcagctgtctcttgaatgcagatggttgcacagtcctgtcttcaaga 2460  
 Db 2353 AGTGATTCACCCAGCTGCTTGTGATCCAGATGTTGATCCAGTTCTGCTTTCAACGA 2412  
 Qy 2461 gaagatttgcagtgatglatgtcaaaaaacgcacaagaacatttcagatgacagt 2520

Db 2413 GAAGGATTGGACGTCAG----- 2430  
 Qy 2521 caattgatttcgttaaaacagaaatacaaaagcatgattagttatagctacag 2580  
 Db 2431 -----ATACCTACAG 2442  
 Qy 2581 actaacatcaatcagtgatgacccagaagcaggttccccaagagatgtgtgct 2640  
 Db 2443 ACTAACTCAATACAGTGATGACCAAGAAAGAGAGTTCTCCAGCAGAGATGTGGTCT 2502  
 Qy 2641 tccctgtctggaagaagcagcgttcgttgaagtcgtcagacccgagttgcagagtg 2700  
 Db 2503 TCCCTGGCTTGAAAGATGAAGTCAAGCTCATTTGAGAGTCTGACACCGAGTTCGGAGGTG 2562  
 Qy 2701 acttgaatggagatattccttccatcgtccagagccgagatcaatcagaagagga 2760  
 Db 2563 ACTTTGAATGGGATATTCCTTCCATGCTCCAGCGCGGATTAATCAGACGAGCGGA 2622  
 Qy 2761 tgcataagagcttcaagagcttccatcgaacaattctatgataaacccggtatagat 2820  
 Db 2623 TGCATGAGAGCTTCAAGCTGACATGCAACAAATCTTATGATAACCCGCGTGAATGAT 2682  
 Qy 2821 gatgtgaagcatgagagaccttggaaagacacacagaagaagttcaagatcagagga 2880  
 Db 2683 GATGATGAAGGCAATGAGACCTTGAAAGAACACAGAAAGTTCAAGATCAGGAGA 2742  
 Qy 2881 ggtctgtatccacagcaggtatcagccttccactctctgtgagagacaatgata 2940  
 Db 2743 GAGTCTGATTCACAGCCAGTGAAGTCACTCCACTCTGTGAGAGCAATGAAATGGA 2802  
 Qy 2941 aacaaagaagaagtgataagactgataaagaagaataaacctggaaagaagaaga 3000  
 Db 2803 AACCAAGAGAAAGTGTATGACTGATAGAAAAAGATTAACATGAGAAAGAAAG 2862  
 Qy 3001 aaagataagataagagaaggaataatgaaagcgaagaaggaatgctgaagctgt 3060  
 Db 2863 AAAGATGAGATAGAGAGAAAGATTAAGAAAGCAAGAAAGGAGATCTGAAGGCTTG 2922  
 Qy 3061 ggaagacatgttcaggttggcgaac 3085  
 Db 2923 GGAGACATGTTCAAGCTTGCCAAAC 2947

RESULT 13  
 LOCUS AK000761 2696 bp mRNA linear PRI 22-FEB-2000  
 DEFINITION Homo sapiens cDNA FLJ20754 fis, clone HEP02246.  
 ACCESSION AK000761  
 VERSION AK000761.1 GI:7021050  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens hepatoma cell\_line: HepG2 cDNA to mRNA, clone\_lib: HEP  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T.,  
 Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Ota, T.,  
 Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
 Nakamura, Y., Isogai, T. and Sugano, S.  
 NEDO human cDNA sequencing project.  
 JOURNAL Unpublished (2000)  
 TITLE 2 (bases 1 to 2696)  
 REFERENCE Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
 Shibahara, T., Tanaka, T. and Nakamura, Y.  
 DIRECT SUBMISSION  
 JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,  
 University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,  
 Minato-ku, Tokyo 108-8639, Japan (E-mail: cdna1@ims.u-tokyo.ac.jp,  
 Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)  
 COMMENT NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan; cDNA full insert  
 sequencing; Research Association for Biotechnology; cDNA library



construction, 5'- & 3'-end one pass sequencing: Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

FEATURES  
source

Location/Qualifiers

1..2696  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEP02246"  
/cell\_line="HepG2"  
/cell\_type="hepatoma"  
/clone\_lib="HEP"  
/note="cloning vector pME18SFL3"  
45..1088  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAA91366.1"  
/db\_xref="gi:7021051"

## CDS

/translation="MKAKKGLKGLDMFRFGKHKRDKIKTKIKIQESTSEER  
IRKOEORIAKTRERERQARENDVAIEDFHRTGCDDELMTGVSSYEGSMALN  
ARPOSREGHMMDALYAQVKPRNSKSPVDSNRPSPNHRIOIRQEBFOAKDEID  
VEDRRRTYSFQPMFNARATQSHSYSEVQMROROEERESSQOARQSSLPQ  
SRKNASSYSDSMEONYPGSGFOSAKENPRYSYOGSNGYLGHGFMARWLETOE  
LIRORRKEOQMKQPSSEGSNDSIKKYQDPSYAPRPFQDVPSPSQVARN  
RLQTPERKRPRTS"

BASE COUNT 703 a 611 c 695 g 687 t  
ORIGIN

Query Match 47.1%; Score 2677.4; DB 9; Length 2696;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2684; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2985 tggaaaagaagaagaatagatagataagagaagaataaataagccaagaagg 3044  
DB 2 TGTAGAACTTCTGAAAAAGATAGATAGAGAGAGATTAATAAAGCCAAAGAGG 61  
QY 3045 aatgctgaaggcttgggagacatgctcaggtttggcaaacatcgaaatgacaagt 3104  
DB 62 AATGCTGAAGGGCTTGGGACACATGTTGCAAGTTGGCAAAACATCGAAAGATGACAAGAT 121  
QY 3105 ttagaaaagggttaataataataacagaatcccttaataatcgaaagagagagatag 3164  
DB 122 TGAAGAAACGGGTAAATAATAATAACAGAAATCCTTTACATCAGAAAGAGAGGATACG 181  
QY 3165 aatgaagcagagcagagagagatcaagccaaactcgagaattagggaaagcaagac 3224  
DB 182 AATGAAGCAGAGAGAGAGAGATTAACCAAACTCGAGATTATAGGAAAGACAGACAG 261  
QY 3225 tcgagagcgtagctatgctgaatcaagaatttcaatcgagaatttggctgtgataga 3284  
DB 242 TCGAGAGCGCTGACTATGCTGAAATTCAGATTTTTCATCGACATTTGCGCTGATGATGA 301  
QY 3285 gttatgtatgtggggaggttcttcttataaggtttccatgagctcaagcttagactta 3344  
DB 302 GTTAATGTAATGGGGAGGTTCTTCTTATGAAGGTTCATGAGCTCTCAAGCTAGACTCA 361  
QY 3345 gagccacagagaagagcatatgatagtatctgttatgcccgaagtcaagagcgcgaa 3404  
DB 362 GAGCCACAGAGAAGGCGATATGATGATGCTTTGATGCCCAAGTATAAGAGCGCGGGA 421  
QY 3405 ttccaaaccttaacctgtatgaacagtaacagatcaactccttagcaatcatgataga 3464  
DB 422 TTCCAAACCTCACTGTGAGACAGTAAAGATCAACCTCTAGCAATCATGATGCGATAC 481  
QY 3465 ggcctcgaagcaagaatttcagaagaagaagaagaatgaagatgtatgaagaatcgttcgg 3524  
DB 482 GCGTCTGAGGACAGAAATTTTCAGACAAAGCAAGATGAAGATGTGAAGATGTGGGCG 541  
QY 3525 gacctaatagttttgaagcaacctgtgcgaacgacgacgacgacgacgacgacgacgac 3584  
DB 542 GACCTATAGTTTGAAGCAACCTGTGGCGAAGCGAGCGCGGAGCAGACAGCGGGGAGCA 601

QY 3585 ctctgtgtccgttggaggttgcagatgcagcgagcgagcgagagagcgagagcttccca 3644  
DB 602 CTCGGTGTCCGTGGAGGTGAGATGCAGCGGACGCGGACAGAGAGAGCGAGAGCTCCCA 661  
QY 3645 ggaagccacagcgcaatgaagctctgtcctgtgcgaagaagaagaataatgcagctcggt 3704  
DB 662 GCAAGCCACAGCGCAATGACGTCTCTGTGCTCGGCAAGAGGAAATAATCCACGCTGGT 721  
QY 3705 ctccacagactcttggagcagaactactccctctggggaaggttcagaagtgcgaaga 3764  
DB 722 CTCGCCAGACTCTTGGAGAGAGAACTACTCCCTGGGGAAAGCTTCCAGAGTCCCAAGA 781  
QY 3765 gaaccacagtlactccaagcttacaaggtctcagaagcagctactctgtggagacatgct 3824  
DB 782 GAACCCAGGTACTCCAGCTACCAAGGCTCCAGAAACGGCTACTGGAGACATGAGCTTT 841  
QY 3825 caacgcagaaggtatgtctgaatactcagaaggtccttcgccaagaagaagaagaagaaga 3884  
DB 842 CAACGCCAGGATCATGTGAAACTCAGAGCTCTTCCAGAGAACAGAGCGGAAAGGA 901  
QY 3885 gcaagcagatgaagaagcagcctccttccgaggggcccagaactatgactcgtataaga 3944  
DB 902 GCAGCAGATGAAGAAGCAGCTCTCTCCGAGGGGCCAGCAACTATGACTGRTAAGA 961  
QY 3945 agtcacagaccacgttaacgcccctcccaaggggcccttcggcaagatgtgccccctc 4004  
DB 962 AGTCACAGACCCCAAGTTACCCCTCTTAAGGGGCCCTTCGGCAAGATGTGCCCTTC 1021  
QY 4005 cccttctcaggttgcgaggttgaacagactcaagctcctcgaaagaagggggccccttcta 4064  
DB 1022 CCTTCTCAGGTGCGAGGTGCAAGGTGAACAGACTTCAGACTCTCGAANAAGGGGCGCTTCTA 1081  
QY 4065 ttccctgaacgcaataaagatgcttcaatgctcgcgcgaataaagaacatttccctatga 4124  
DB 1082 TTCTCAGACGACGAATAATACAGATGCTTCATGTCGCCATTAATAAGCATTTTCRTATGA 1141  
QY 4125 agactgtatcttggagagtttctttaaaccctgcagtgatctatggaatcttcgttgt 4184  
DB 1142 AGACTGTATTTTGGGAGGTGTTTAAACCTCGATGCTGATGAGATTTTCCTGTTGT 1201  
QY 4185 tggatcagagtcctttaaaggtgtgtggaagaagaatggaagccttaagttcttggcca 4244  
DB 1202 TGGTATCAGTCCCTTTAAAGGGGTGATGAGCAAAATAAGGAAGGCTTAATGCTTTGCCA 1261  
QY 4245 ctatgtcacaagtgctcgttttcaatggaagattcccaacctgtgaacaacatcgttgt 4304  
DB 1262 CTATGCTCAAGTGTGCTGTTCAATGGAAGATTTCCACACCTGTGACAAATCATCTGTTGG 1321  
QY 4305 aggtgtcataatgctctgcctcctcccaagtaacgaagatctcgccctactcaatgagt 4364  
DB 1322 AGGTGTCAATATGCTGCGCCCTCTCCACAGTACCAAGAAATCTCGCCCTACATCATAGT 1381  
QY 4365 tgcgcgaggttgtgtgaacatccctgcacacacttgcagatgaatgaactgaagt 4424  
DB 1382 TGTCCGGGGTGTGTTAATCATCTCGACACACTTGCATGACAAATTAACCTGAAGTG 1441  
QY 4425 gaagatgaagtcgagccctgttctccctcaaggttctcctaagctatgagatgacatct 4484  
DB 1442 GAGGATGAGCTGGGCGCCTGTTCTCCCTCTTAAGTTCTCTTACCTATGAGATGACATCTT 1501  
QY 4485 agtctctgttgaagaaagtggcgagacataacaaatgtgggcttctcgttactc 4544  
DB 1502 AGTCTCTGGTGGAGGAAATGGGCGATATACACAAATAATTTGGGCGCTTCTGTACTTC 1561  
QY 4545 acagcaagccatgttgcactcttgacatcactgtgtgttctccttctcctcagct 4604  
DB 1562 ACAGCACAGCATTTGTGCTACTTTGTCAATCACTGTGTTTCTCTTCTCTCAAGT 1621  
QY 4605 ctcttgaacgggaagtcggtcatcctatatcagaagctaaagcatagttccaaactgtt 4664  
DB 1622 CTTTGTGACGGAGAGTGGGTCAATCCATTAACAGAAAGCTTAAGCAATGTCACAACTGTT 1681  
QY 4665 tggtaacaatgggggtcccttcttgaactgccttataatgaactaacaataaagtat 4724

[illegible]

VERSION	BD004905.1	GI:18632866
KEYWORDS	JP 03074935-T/2.	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens.	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 2718) Ota,T., Isoqali,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Nakamatsu,A., Nagai,K., Otsuki,T., Funahashi,S., Seno,C. and Nezu,J.	
TITLE	Novel genes encoding protein kinase or protein phosphatase Patent: JP 03074935-T 2 30 JAN-2001.	
JOURNAL	HELIX RESEARCH INSTITUTE, TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KEIJI HAYASHI, KAORU SAITO, JUNICHI YAMAMOTO, SHIZUKO ISHII, OMOKASU SUGIYAMA, AI WAKAMATSU, KEIICHI NAGAI, TETSUJI OTSUKI, SHINICHI FUNAHASHI, HIROKI SENO, JUNICHI NEZU	
COMMENT	OS Homo sapiens (human) PN JP 03074935-T/2 PD 30-JAN-2001 PE 28-JUL-2000 JP 2000005061 PR 29-JUL-1999 JP 99P 248036, 11-JAN-2000 JP 00P 118776 PR 02-MAY-2000 JP 00P 183767, 09-JUN-2000 JP 00P 241899 PR 18-OCT-1999 US 60/159590, 17-FEB-2000 US 60/183322 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KEIJI HAYASHI, PI KAORU SAITO, PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU, PI KEIICHI NAGAI, TETSUJI OTSUKI, SHINICHI FUNAHASHI, CHIHIKI SENO, PI JUNICHI NEZU PC C12N1/5, C12N9/12, C12N9/16, C12N1/15, C12N1/19, C12N1/21, PC C12N5/10, C07K16/40, PC C12Q1/68, C12Q1/42, C12Q1/48 CC FT FH Key Location/Qualifiers CDS (33)..(2627).	
FEATURES	source 1..2718 /organism="Homo sapiens" /db_xref="taxon:9606"	
BASE COUNT	840 a 586 c 688 g 604 t	
ORIGIN		
Query Match	43.5%; Score 2476.8; DB 6; Length 2718;	
Best Local Similarity	96.0%; Pred. No. 0;	
Matches 2603; Conservative	0; Mismatches 7; Indels 102; Gaps 2.	
QY 374	ttgaggtcaacaccccttcagctcttggaagcaataatgctcttcattatgtttggaagcagtatgt	433
DB 1	TTGAGGTCAACACCTTCAGCTTCCTTCGACGAAATATCCCTCTTCATATGTCACGACGATG	60
QY 434	accgagctcttaattgagcctctcaccctctgctcaatgcatgatgaataattttcctctgaagac	493
DB 61	ACCGAGCTCTAATTGGCGCTCTCCACTTCGTGATGATGATGTAATTTTCTCTGAAAGAC	120
QY 494	cttcaagaagaaaaatccacacagctgtgtcaacaacagctgtgcttcctcaagcaagacacty	553
DB 121	CTTCAAGAGAAAAATCCACACGCTGTGTCAACACGCTGCTTCCTTCACAGCAACACAG	180
QY 554	ctggagagctcttaaacaccttcgacacaggaagaagatgaaaaaactacacagaagcctcccgcg	613
DB 181	CTGGAGAGCTCTTAAGCCTTCGACAGGAGAAAGATGAAAACTACACAGCCTCCCGCGG	240
QY 614	atactagtaactgtgtcttaacccaattcagagagacaatgtcgtctgctctgtagtgcga	673
DB 241	ATACTAGTAACTGTGTCTTACCAATTTCAAGAGACAACTGCTGCTGCTGTGAGTGCA	300
QY 674	gtcaccacaatggtgtggcagaatgtgtgtggagaacaagaagaagatgaatgtgtggacagaag	733
DB 301	GTCACCAATGTGTGGGCAAGTGCGAGGAGAAACAAGAAAGATGAGTGTGGAGACAGAAG	360
QY 734	aggataaacagctgtgttggaacctgtttgacatgttggacatgctggaacgggtttggaacatacca	793
DB 361	AGGATTAACAGCTGTGTGAACCTGTTGGACATGCTGACACAGCGGTTTGGAGCATATACCA	420

OY	794	accttctctgatatgataatgtaagcttcgtatgaagtcctccaaagatgtaggagcctctg	853
Db	421	ACTTTTCTCTGATGATATATGATTAACCTCGTGAAGTCCCAACAGATGAAGGCTCTGG	480
OY	854	gaatccatgtaatgaccttcaatgctcgaggggagaaacctggggtatataa	913
Db	481	GAATCCATGTAATGACCTTTTACGTGCTGAGGGCGCAAAACCTGGGGTATTAATGA	540
OY	914	gattgagaaagglygtaagcctgcaatgaaatcttctcgtatgaatgatcgatg	973
Db	541	GATTGAGAAAGGTGGTAAAGCTGAACATGAATAATCTTTTGTGATGATGATTCGAT	600
OY	974	tcaggaatlaatgaltgycgaccttcgaatagaagatttgaaacagacacatatgttc	1033
Db	601	TCAGGATTAATTAATGATGGCGACCTTCGAATAAGAAATTTGAACAAACATATATG	660
OY	1034	ggcaagcgatcgtaaacccatcatattggttcacatggttctcgcagcaataaagac	1093
Db	661	GCCAAGCGCATGCGTACACCCATCATTTTGGTTCCATGTGGTCTTCACAAATTAAG	720
OY	1094	agatgcaacatctcccaagtgaaagaacaaatctacttcaagccggtttagccgtg	1153
Db	721	AGTATGAACAACATCTTCCAAAAGGAAABAACATTAATTAATCAAGCCGTTTAGCCT	780
OY	1154	acagccagatattgacaacagagatgtgacagtgcagggctcacaggtgcagaag	1213
Db	781	ACAGCCAGTATTAATTTGACAAACAGAGAGGTGTAACAGTGACAGGCTTCACACG	840
OY	1214	cacccgacatgaaacaccccgctcgagcaatgactctcaactcaagactactatagc	1273
Db	841	CACCCCGACATGAACACCCCGCTGACCAATGAGCTCTCACTCAAGATACCTCAT	900
OY	1274	caacccctcgggaaaaacacacaccccgcccgccagccttcgagcactagaatlatia	1333
Db	901	CACACCCCTCGGAAAAACACACATCCGCTCCAGCCTCGGACCTCAGAAATGATTTA	960
OY	1334	cgactgtaagcaglygtataacacccaaaaataggaagaagccttaataaccagctta	1393
Db	961	CGACTGTAAACATAGTGTATTAACACCAAAAAATAGGCAAGAGGCTTAATATCCAG	1020
OY	1394	agaaagttacaagaagtttgggalttaagatcaattccagagatgtaacaaatagtgtc	1453
Db	1021	AGAAAGGTACAGAAAGGTTTGGGATTCAGATCATCTTCCAGAAATGTAACAAT	1080
OY	1454	cagctccaactctatgtaaaaaacatctcccccggggggcgccactlaagatggccag	1513
Db	1081	CAGCTCCAAATCTATGTAAAAAACATTTCTCCCCGGGGGGGCCATTGAGATGGCG	1140
OY	1514	ttaagcagaagacagactatagaggtaaatgtagatltagttgggcaatcccaag	1573
Db	1141	TTAAGCAGAGGACAGACTTATAGAGGTAAATGGAGTAGATTAGTGGCAAAATCCAG	1200
OY	1574	aggaagtgttgcgtgttgagaagaccaaagatvgaaagacvtgagccctcgtgct	1633
Db	1201	AGGAAGTGTGTTTGGCTGTGTAGAAAGCACCAAGATGGAAGAACTGTGAGCC	1260
OY	1634	ttcgccaggaagaagcccttcaccaccaagggaactaaatcaaaagccaagccaatg	1693
Db	1261	TTCCGCGAGGAAGACGCTTCCACCCCAAGGAACTGAATCAAGCCAAAGCCAGATG	1320
OY	1694	tctcaaaagaaaacgaagacgaagaatgtagatatttcttcaacctgtagccaccag	1753
Db	1321	TTCCAAAAGAAACGAAGACGAAGATGAGAGATTTGTCTTACACTGANTGGCAC	1380
OY	1754	aattctcgacaatttgaagttcccaactaaatgattcaggaatctgcaggtcttggtc	1813
Db	1381	AATTTCTGACATTTTGAAGTCCCACTTAATGATTCAGAGATTCACAGCCTTG	1440
OY	1814	tcaaagttaaacccgttcaaaagagaaacacagcaatltgggaatctcttgaaatcc	1873
Db	1441	TCAAAAGTAACCGGTCAAAAAGAAACACGACAGATTTGGGAATCTTTGTCAAGT	1500
OY	1874	ttaatgaggaagcagcatctaagaatggaaggtcttcgggtgaatgatacactgata	1933

Db	1501	TTAATGAGGAGCAGCATCTCTAAAGATGGGAAGGCTTCGGGTGTAATATCACTGATAGCAG		1560
Qy	1934	taaatlgagaatccctcttggcgagaacaaacgaatgcctatgnaacccttaagaagt		1993
Db	1561	TAAATGGGAATCCCTCTTGGCAAGACAAACCAATGCCATGGAAACCTTAAGAAAGT		1620
Qy	1994	ctatgtctactgtaaggaataaacaaggaatgatccagcttatctgtgcaaggaataa		2053
Db	1621	CTATGTCTACTGAAGGCATAAATAAGAGAAATGATCCAGTTATTTGTTCAGAGGAATAA		1680
Qy	2054	gcaagtgcaatgagctgaagtcacctcgggagcccccggaccctgaagctgcccattgaa		2113
Db	1681	GCAAGTGCAAATGACCTCAAGTCAACCTGGGAGCCCCCTTGACCTGACGTGCCCATTTGAA		1740
Qy	2114	cagcgcttgatgataagagaacgaagaattcccatccctcctaagtgaggatttgaaggc		2173
Db	1741	CAGCGCTTGATGATAGAGAAACGAAGAAATTTCCATTCCCTTACAGTGGGATTGAGGGC		1800
Qy	2174	ttgatgaatcccgacagaatactgcctccctcagtagaataatgggtgaatcgaataat		2233
Db	1801	TTGATGATCCGCCAGCAAAATGCTGCCCTCAGTAGAATAAG-----GGTAAT		1861
Qy	2234	accagctgccccctacagtgataatgccccagaatgacatgcatatagaagaatgaca		2293
Db	1852	ACCAAGCTCCCTCCCTACAGTAAATATGCCCCAAGATGACACTGCAATTTAGAAATGACA		1911
Qy	2294	ggttgcaagtgcttccctccacatctctcagacagtcctccttccagctcccatgtagtg		2353
Db	1912	GGTTGCCAGTCTTCTCCACATCTCTGACCAAGTCTCTTCCAGCTCCCTCATGTAAATG		1971
Qy	2354	tgaggcttgtagcgcagagatgctgtagtacttggcgcaagctgcaatcagtgataacgcg		2413
Db	1972	TGGGTTGTGTAGCGGAGATGCTGTGTAATCTTGGGCCAAGGCTGCATCAATGATTAGCGG		2031
Qy	2414	actgctcttgtagtcscagatgttgatccagttcttgctttcaacgagaagatttggac		2473
Db	2032	ACTGCTCTTTGAGTCCAGAGTGTGATGCCAATGCTTCTTCAACAGAGAAATTTGGAC		2091
Qy	2474	gtcagagatgtgcaagaaaacgcacaagaatcttcagatgcagctcaattgattctcg		2533
Db	2092	GTTCAA-----		2097
Qy	2534	ttaaacaacgaataaataaagaatgatttagtatagtctacagacttaactcaata		2593
Db	2098	-----CTGACGAGACTTAATCAATA		2118
Qy	2594	cagtgatgaccaaagaacaggtctctccacgagaatgttggtctctccctggtctga		2653
Db	2119	CAGTGAATGACCAAGAAAGAGGTTCTCTCCAGAGATGTGGGTCCTTCCGCGGTGCA		2178
Qy	2654	agaagtcagatcgttgtagagatgtctgacacgcgcagttgcgagtgacttgaatggg		2713
Db	2179	AGAACTCAAGCTCAATTTGAGAGTCTTGACAGCCCAATGCCACAGGTGACTTTGAATGGG		2238
Qy	2714	atatcccttcacatgcgcacgcgcgcgcgagataatcagagcgagggatgacatgagact		2773
Db	2239	ATATTCTTTCCATTCGTCCACGGCCGGGTAATTCAGAGGCAAGGGATTCATAGAGACT		2288
Qy	2774	tcagagctgcatcgacaatatctatgataaacccgcggtlagaatgataatgataagga		2833
Db	2299	TCAGAGCTGCATCGACAATACTTATGATTAACCCCGGTAAATGATGATGATGAAGCA		2358
Qy	2834	tgagagacttggaagaagacacacgaagaagatctcaagatccaggagagagatctgatacca		2893
Db	2359	TGGAACTCTTGGAAAGAGACACAGAAATAAATTCAAGATTCAGGAGAGCTCTGTATCA		2418
Qy	2894	cagcagatgacagccttccacatctcttgtagagacaataatgataagaaacgaagaag		2953
Db	2419	CAGCGAGATCAACCTTCCACCTCTCTGAGAGACAAATGAATGGAACCAAGGAAG		2478
Qy	2954	gtgataagactgatagaanaaagataaaactgnaaagaanaagaagataagata		3013

QY	3014	agagaagataaataatgaaagccaaagaaggaaggaatgcctgaagggcttggagacatgttca	3073
Db	2539	AGGAGAAAGATATAAATGAAAGCCAAAGAGGAAATGCTGAAGGGCTTGAGACATGTTCA	2598
OY	3074	gatttggaacac	3085
Db	2599	GCCTTGCCAAAC	2610
RESULT	15		
AC027735			
LOCUS			
DEFINITION			
AC027735			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
Source			
CDS			

		EGFRRITDRERIKLNTVDDQKXGSPSRDVGPSLGIKSSSLISQIWALEVLTNGDIPIPH	
		RPRRIITRGRGCGNSFPAALIDKSDYDRAVDDGDMLEEDTPESSRSSEYSTAS	
		DQPSHSLEROMNGQERGDTRDKRDKTGKEKKRDKRDKRDKMKAUKMLKGLGDMS	
		LAKLPEPR*	
BASE COUNT	840 a 586 c 688 g 604 t		
ORIGIN			
Query Match	43.5%;	Score 2476.8;	DB 9; Length 2718;
Best Local Similarity	96.0%;	Pred. No. 0;	
Matches 2603; Conservative	0; Mismatches	7; Indels	102; Gaps 2;
Oy	374	ttgaggtcacaccttcagtccttcgcagcaaatatgctcttcataigtctgaagcagtagtg	433
Db	1	TTGAGGTCAACACCTTCAGTCCCTTGAGCAAAATATGCTCTTCATGTTCGACGCACTAGTG	60
Oy	434	accagcctctaattggtcctctccactcttgtaagtgatgaatatttttccttcgaagc	493
Db	61	ACCCAGCCTTAATTTGGCCTCTCCACTTCCTGTCAGATAGTAATTTTCTCTGAAAGAC	120
Oy	494	cttcaaggaanaatcccaacgcgtgtcaacaacagcttggtcttcctcaagcagaacatgy	553
Db	121	CTTCAAGGAAAAATCCACAGCGTGTCAACAACAGCTGGCTTCCTCAAGCAAGAACACTG	180
Oy	554	ctggagagtccttaaaacacctcgacaggaagaagaatgaaatactacaagaagctccgcggyg	613
Db	181	CTGGGAGTCTCTAAACCTTCGACAGGAAGAAAGATGAAAACTACAGAAACCCCTCCGCGG	240
Oy	614	atacagtaaacggtggtctaaccaattcaagagacatgctgcgcgtctcgaagtcca	673
Db	241	ATACTAGTAAGTACTGGTCTTAACCAATTTACAGAGACATCTCTGCTGCTCTGAGTGCCA	300
Oy	674	gtcaccocaaatggttgggcaagtgctgtagaanaacaagacaggaatgagatgagacagaag	733
Db	301	GTACACCCATGATGGGCAAGTGGCAGGAGAAACAAGAACAGATGAGATGGGACAGAG	360
Oy	734	agagataacgctggtttgagacctgttgagcatgctggaacaggggtttgagcatabacca	793
Db	361	AGGATATACGTGTGTGTGAACCTGTGTGACATGTCTACACGGGTTTGGACATATACCA	420
Oy	794	actttctctatgatataatgtaaaagcttgtagaagtcctcccaacgaatgagaggtcctcgg	853
Db	421	ACTTTTCTCTGGATGTATGTATAAAGCTGTGAAGTCCCCAACGATGAGGGCCTCTGG	480
Oy	854	gaatcacatgtagtgccttcagtgctcgagggcgagcaaaccttggtattagtaaac	913
Db	481	GATCATATTAATGCTCTTTCAGTGTGCGGGCGGAGAACCTCGGGGTTATTAATGAAC	540
Oy	914	gatttgagagaagtggtgtaagctgtaacatgaaatctcttcctgtgagaatgattgcatgy	973
Db	541	GATTGAGAAAGGTGTTAAGCTGAACATGAAGAAATCTTTTGGTAGAATGATTCGATTG	600
Oy	974	tcaagattaatgtagtgagaccttcgaatatagaagatttgaacaagcaacaatatgttcc	1033
Db	601	TCAGGATTTATGATGTGCGACCTTCGAAATAGAAAGATTGTAAACAAGCAACAACATATTTTC	660
Oy	1034	ggcaagccatgctgtacaccatcatttggttccatgttggttccctgagcaaatgaaagag	1093
Db	661	GCCAAACCATGCGTAAACCCATCATTTGGTTCATGTGGTCTCTGACGAAATAAAGAGC	720
Oy	1094	agtatgaacaaactacccaagttagaagaacaatactactatcaagcgttttagccctgy	1153
Db	721	AGTATGAACAAACTATCCCAAAGTGAAGAAACATTAATTAATTCMAACCGTTTAAAGCCATG	780
Oy	1154	acagcgagatataatggcaaacagagtggtataaacgttgcaaggtcttcaacagtgccaagag	1213
Db	781	ACAGCAGATATATTGAACAAAGAGTGTGACAGTCAAGGCTTTCACACGGTGCACAGAGC	840
Oy	1214	caaccgcagtagaacacacccgctcgagcaagatagacatcctaactcaagaactacatcatagc	1273
Db	841	CACCCCGAGACTGAACCAACCGCCTGAGACAGATAGTACTCTCACTCAAAAGTACCTCATAGCGC	900
Oy	1274	cacacccctcgaggaaaacacacatccgctctcagctccgcgacactcgaatagtatttagta	1333

Db	901	CACACCCCTGGGAAACCAACACATCCGCTCCACCTCGGACCTCGACCTAGAAATGATTTAGTA	960
Qy	1334	cgactgtgaagcagtggttataaacccaanaaataggcaagagcttaatatccagctta	139
Db	961	CGACTGTGAAGCAGTGGTTATTAACCCAAAAAATAGGCAAGAGGCTTAATATCCACTTA	1022
Qy	1394	agaaaggtacagaaggtcttgagatccagactcaactccagaagctgtlaacataagtgagct	145
Db	1021	AGAAAGGTACGAAGGTTTGGGATTCAGCTACACTTCCACAGATGTTACATATAGTGGCT	108
Qy	1454	cagctccaatcatalgtgaaaaaatctcccccgggggcygcacatcaagatgycgcag	151
Db	1081	CAGCTCCAAATATATGTAAAAAATCTTCCCGGGGGGGCGGCATTCAGATAGCGCGAC	114
Qy	1514	ttaagcagggagacagacttatagagttaatggagtgataggtttaggtgggaaatcccaag	157
Db	1141	TTAAAGCGAGGACACACTTATGAAGTTAAATGGATGATGATTAATAGTGGCAAAATCCAAAG	120
Qy	1574	aggaagttgttcgtctgtgagaagcccaagatgagaagaaactgtgaacctctgtct	163
Db	1201	AGGAAGTTGTTTCCCTGTGTGAAGACCAACAAGATGGAAGAACTGTGAGCTTCTGTGTCT	126
Qy	1634	ctgcgccaggaagagccttccacccaagggaactgtaatcgaagcccaagccagatgcaga	169
Db	1261	TTTCGCCAGGAAGACGCTTCCACCCAAAGGAACTGAATGAGAGCCAAAGCCAAATGCGAG	132
Qy	1694	ttccaaaagaacgaaacgaaagataggaatattgtcttaaacctgatacctgatacgaaggt	175
Db	1321	TTCCAAAAGAAACCAAAAGCAAAAGATATAGATATGTCTTACACCTGATGTGCACCAAGG	138
Qy	1734	aattctgaacattgaagctccaccataagatcgaatcgaagccttgtgtcagtg	181
Db	1381	AATTTCTGACATTTTGAAGTCTCCACTTGTAGTATTCAGCATCTGCAAGGCTTGTGTCAAGTG	144
Qy	1814	tcaaaagtaaccggttcaaaaagaagacacgacgaagtttgggaactctgttaacttcactta	187
Db	1441	TCAAAAGTAAACCGGTCAAAAGAGAACCAACGACGAGTTTGGGAATCTTTGTGAAGCCACTTA	150
Qy	1874	ttaatlgagaagcagcatctaaagaatlgaaaggtctcggtgtaatlgatacactgatacag	193
Db	1501	TTAATGGAAGGAGCAGCATCTTAAAGATGGAAGGCTTGGGGTAATGATCAATGATATGAC	156
Qy	1934	taaatlgagaatccctctgttgggcaagacaaacaaagtgcacatgaaacccctaagaagct	199
Db	1561	TAAATGGAGAAATCCCTGTGGGGCAAGCAAAACCAAGATGCGATGGAACCCCTAAGAAAGGT	162
Qy	1994	ctatgtctactgaaagcgaataaagagaatgatacgaactatattgttgcagaagagaaata	205
Db	1621	CTATGTCTACTGAAAGGCAATTAACAGAGATGATGATCCAGCTTATTTGTTGCAAAGGNAATTA	168
Qy	2054	gcaagtgcaatgagctgaagtcacacttggagagcccccccttggacctgagcttgcacattgaa	211
Db	1681	GCAAGTGCAAATGACACTGAAGTACACTGTGGAGCCCCCTTGACACTGACACTGCCATTGAAA	174
Qy	2114	cagcgttggatgatatagaagaagaagaatttccacttccctcctcagtgaggtattgaggggc	217
Db	1741	CAGCCTTGGATGATGAACAACGAAGAATTTTCCATTCCTCTACAGTGGGAAATGAGGGGC	180
Qy	2174	ctgatlgaatctgcacagcagaatctgctgcctcagtaagataatlygtgtgagctcagtleaat	223
Db	1801	TTGATGATTCGCCACGACGAATAATCTGGCCTCACTAGAGATATG-----GRTAAT	185
Qy	2234	accagctgtccctccacagtgtaatttgcgcccaagatgacactgtcatatatagaagatgaca	229
Db	1852	ACCAAGCTTCTCCCTTACATGATGATATGCCCCCAAGATGACACTGCTATATATGAAAGATACA	190
Qy	2294	ggttgcaggtcttccctccacatctctcgaacagctcctcttccagcttcccaatgatatg	235
Db	1912	GGTTGCCAGTGCTTCTCCACATCTCTCTGTGACAGTCTCTTCCAGCTCCCATGATGATG	197
Qy	2354	tggggcttgtgacggcagatgctgtgtaacttgggccaagagctgcaatcagatgattcagcgg	241

Db	1972	TGGGGTTTGGAGCGGAGAGATGCTGATCTTGGGCGCAAGGCTGCATAGTATTGACGG	2031
OY	2414	acgcctctttgagtcacagatgltgtatccagcttccttgcttccacgagaagatttgac	2473
Db	2032	ACTGCTCTTTGAGTCGCGAGATGTGATTCACGGTTCTTCTTTCAAGAAAGATTGGAC	2091
OY	2474	gtcagagtatgtcagaaaaaacgcacaagaatlttcgatagtccagtcgaatttgatttcg	2533
Db	2092	GTTCAGA-----	2097
OY	2534	ttaaacacgaaatacctaaaagatgtattagatagctgacgagactaaactcaata	2593
Db	2098	-----CTACGACACTAACTCAATA	2118
OY	2594	cagctgatactgacccagaagacagctgtctccacgacagatgtggtgtcctccctggtctga	2653
Db	2119	CAGTGGATGACACAGAAAGCAGGTTCTTCCACGACGATGTGGGTCTTCCCGGGTCTGA	2178
OY	2654	agaaagtcgaagctglttggaagatctgcagaacgcagctgtgcgaagtgacttgaattggg	2713
Db	2179	AGAAGTCAAGGTCATTTGGAGAGTCTGCAGACCGCATTTCCAGGCTGACTTTGAAATGGGG	2238
OY	2714	atattcccttccatcgcgcacgcgcgcggaatacgaagcgagggaatgcgaatgaagact	2773
Db	2239	ATATTCTTTTCATCTGCTCCAGCGCCGGGTAATATCAGAGCGCAGGGGATGCAATTGAGACT	2298
OY	2774	tcaagactgtccatctgcacaaaactcttatgatataacccgcggtlaagtatgtatgaatgaagca	2833
Db	2299	TCAGAGCTGCATCGACCAAACTTATGATPATAACCCCGGAGATGATGATGATGAAGCA	2358
OY	2834	ttggaagcctttggaagaagaacacacgaagaagaagttccaagatcgaaggagaagcttctatcca	2893
Db	2359	TGGAGACCTTTGGAAGAAGACACAGAAATAATTCAGATCTACGGAGAGAGTCTGTATCA	2418
OY	2894	cagccagtgatcagccttcccaactctcttgtagagacaatatgatatgaaaccaaaggaagaag	2953
Db	2419	CAGCAGTGATCAACCTTCCCACTCTCTGAGAGACAAATGAAATGAAACCAAGAAAG	2478
OY	2954	gtgaataagactgatagaaaaaaggatataaacttgaaagaagaagaagaagaataagaata	3013
Db	2479	GTAATTAAGACTGATAGAAAATAAAGGATPATAACTGAAAAGAAAAGAAAGATPAGAGTA	2538
OY	3014	aggaagaagataaaaaatgaagaagccaagaagaagatgcttgaaggccttggagacatgttca	3073
Db	2539	AGGGAAGGATTAATAATGAAAAGCCAGAAAGGATGCTGAAAGGGCTTGGAGACATGTTCA	2598
OY	3074	ggtttggcaac 3085	
Db	2599	GCCTTGGCCAAAC 2610	

Search completed: July 24, 2002, 11:12:39  
Job time: 16331 sec



Thu Jul 25 08:38:18 2002

OM of: US-09-757-781-2 to: GenEmbl:\* out\_format : pfs  
Date: Jul 24, 2002 6:10 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=frame+2, model -DEV=x1h  
-O=/cgn2.1/USPTO.spool/US09757781/runat.22072002.154317.18590/app\_query.fasta.1.1434  
-DB=GenEmbl -OPMT=fastap -SUFFIX=olig.rge -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPECL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000  
-DELLOP=6.000 -DELEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000  
-DELLOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=olig  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=quality  
-THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=fts -NORM=ext  
-HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09757781.ecgnl.1.5806 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -NO\_XLPHY -WAIT -THREADS=1

## Search information block:

Query: US-09-757-781-2  
Query length: 1356  
Database: GenEmbl:\*  
Database sequences: 1797656  
Database length: 187333701  
Search time (sec): 3888.640000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

## score\_list:

Sequence	Strid	Olig	EScore	Len	Documentation
gb.pr.AF67002	+	1356.00	25804.76	0.0	4071   AF467002 Homo sapiens partitl
gb.pr.AF67003	+	1113.00	21171.89	0.0	4062   AF467003 Homo sapiens partitl
gb.pr.AF196185	+	1113.00	21169.30	0.0	5958   AF196185 Homo sapiens atypica
gb.pr.AF467004	+	1025.00	19494.31	0.0	3960   AF467004 Homo sapiens partitl
gb.pr.AF196186	+	781.00	14843.26	0.0	3470   AF196186 Homo sapiens atypica
gb.pr.AK024668	+	614.00	11662.69	0.0	2116   AK024668 Homo sapiens cDNA: H
gb.pr.AE332552	+	556.00	10554.61	0.0	2867   AE332552 Homo sapiens atypica
gb.pr.AE332553	+	556.00	10552.90	0.0	3822   AE332553 Homo sapiens atypica
gb.pr.AE67006	+	556.00	10552.70	0.0	3933   AF467006 Homo sapiens partitl
gb.pr.AE52293	+	500.00	9485.27	0.0	3801   AE52293 Homo sapiens partitl
gb.pr.BC011711	+	497.00	9433.97	0.0	1592   BC011711 Homo sapiens clone M
gb.pr.AP177228	+	427.00	8097.84	0.0	2004   AP177228 Homo sapiens cDNA: H
gb.pr.BD004905	+	420.00	7962.32	0.0	2718   BD004905 Novel genes encoding
gb.pr.AK027735	+	356.00	6742.19	0.0	2696   AK027735 Homo sapiens cDNA: FL
gb.pr.AK000761	+	318.00	6015.50	0.0	3735   AF467005 Homo sapiens partitl
gb.pr.AK067005	+	255.00	4819.43	4.4e-260	1772   AK067005 Homo sapiens cDNA: FL
gb.pr.AK025892	+	131.00	2464.99	6.1e-129	425   AK025892 Homo sapiens cDNA: FL
gb.pr.AK072873	+	131.00	2427.61	7.4e-127	106198   AL160409 Human DNA sequence
gb.pr.AK025823	+	131.00	2425.80	9.4e-127	138688   AC025823 Homo sapiens cDNA: FL
gb.pr.AK025823	+	131.00	2423.44	1.3e-126	196374   AL591464 Homo sapiens cDNA: FL
gb.pr.AK025823	+	129.00	2389.59	9.7e-125	104357   AL392123 Human DNA sequence
gb.pr.AK025823	+	102.00	1885.29	1.2e-96	22279   U822190 Homo sapiens chromosome
gb.pr.AK025823	+	89.00	1621.95	3.3e-84	597   AX187060 Sequence 2755 from Pat
gb.pr.AK025823	+	88.00	1627.84	2.6e-82	5500   AB005549 Rattus norvegicus mRN
gb.pr.AK025823	+	84.00	1569.33	4.7e-79	400   AX071341 Sequence 1813 from Pat
gb.pr.AK025823	+	81.00	1471.27	1.4e-73	165146   AC011644 Homo sapiens clone
gb.pr.AK025823	+	81.00	1471.27	1.4e-73	167163   AC013705 Homo sapiens clone
gb.pr.AK025823	+	81.00	1471.28	1.4e-73	169337   AL356462 Homo sapiens cDNA: FL
gb.pr.AK025823	+	81.00	1470.29	1.4e-73	193131   AL390766 Human DNA sequence
gb.pr.AK025823	+	78.00	1452.91	1.5e-72	540   AX321805 Sequence 336 from Pat
gb.pr.AK025823	+	77.00	1420.00	9.9e-71	4167   AT026057 Mus musculus PAR-3 18
gb.pr.AK025823	+	60.00	1073.74	1.9e-51	109816   AL360233 Human DNA sequence
gb.pr.AK025823	+	60.00	1072.18	2.3e-51	138322   AC012241 Homo sapiens cDNA: FL
gb.pr.AK025823	+	60.00	1069.81	3.2e-51	196259   AL392171 Homo sapiens cDNA: FL
gb.pr.AK025823	+	52.00	918.33	8.7e-43	168213   AC103175 Rattus norvegicus c
gb.pr.AK025823	+	44.00	808.55	1.1e-36	305   AX187694 Sequence 3389 from Pat
gb.pr.AK025823	+	44.00	808.47	1.1e-36	309   AX186627 Sequence 2322 from Pat

gb.pr.AX184822	+	42.00	769.71	1.6e-34	339   AX184822 Sequence 517 from P
gb.pr.AX184822	-	40.00	689.31	4.9e-30	173987   AL450337 Human DNA sequen
gb.pr.AX184822	-	38.00	651.45	6.3e-28	167163   AC013705 Homo sapiens cto
gb.pr.AX184822	+	34.00	614.50	7.3e-26	504   AX184822 Sequence 1329 from
gb.pr.AX184822	-	34.00	574.83	1.2e-23	176405   AX184822 Human DNA sequen
gb.pr.AX184822	+	33.00	562.01	6.1e-23	70087   AC105981 Mus musculus clon
gb.pr.AX184822	-	28.00	466.69	1.2e-17	70087   AC105981 Mus musculus clon

## seq\_name: gb.pr.AF467002

## seq\_documentation\_block:

LOCUS AF467002 4071 bp mRNA linear PRI 07-FEB-2002  
DEFINITION Homo sapiens partitioning-defective 3 protein splice variant a  
(PAR3) mRNA, complete cds, alternatively spliced.  
ACCESSION AF467002  
VERSION AF467002.1 GI:18568343

## SOURCE

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

gene

CDS

misc-feature

/gene="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

/note="PAR3"

```

misc_feature      /note="Region: PDZ domain"
                  /gene="PARD3"
                  /note="Region: PDZ domain"
misc_feature      /note="PARD3"
                  /gene="PARD3"
BASE COUNT       1220 a 923 c 1079 g 849 t
ORIGIN

```

## alignment\_scores:

```

Quality: 1356.00      Length: 1356
Ratio: 1.000          Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x AF467002 ..

```

Align seg 1/1 to: AF467002 from: 1 to: 4071

```

1 MetLysValThrValCysPheGlyArgThrArgValValValProCysgl 17
  |||||
1 ATGAAGTGAACCGTGTCTCGACGAGACCCGGTGGTGGTCCCGCGCGG 50
17 YASGLYHISMetLysValPheSerLeuIleGlnAlaValThrArgT 34
  |||||
51 GGAGGCGACATGAAAGTTTCAGCTCATTCACGCGCGGTGACCCGCT 100
34 YHArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnAlaHsArg 50
101 ACCGGAGGCCATCGCCAGAGATCCAACTACTGCTACAGTGCATCGC 150
51 LeuGlnHsArgLysAspGlyIleLeuAspLeuAspAspIleLeuCysas 67
151 TTGGACATGAGATGAGAGAAATGACTGACCTGATGACATCTTTGTGA 200
67 PValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84
201 TGTAGCAGACGATAAAGACAGACTGTGACAGGTGTGATGACAGAGATC 250
84 rHisHsIleGlyLysAspGlyThrSerAlaSerSerThrGlyHsArg 100
251 CACATACAGGAGGTGATGGCCAGCTGCAAGTCCACGCGGTACCCAGGC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG 117
301 CCACAGATATTTGTGTAGTGGCTTGCGACCAACATGCTCAGCTTTTCA 350
117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
351 GCCTTACCAAGCAACAGTGAATTTGAGTCAACCTTCAGTCTTCGAG 400
134 LAsnMetProLeuHsValArgArgSerSerAspProAlaLeuIleGly 150
401 CAATATGCTCTTCATGTTGAGACGATGACCCAGCTCTATTTGCG 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
451 CTCCTCACTCTGTCAGTGAATGTAATTTTCTCTGAAGGCGCTCAAG 500
167 gLysAsnProThrArgTyrSerThrThrAlaGlyPheLeuLysGlnAsnT 184
501 GAAAAATCCACACGCTGCTCAACACAGCTGGCTTCTTCAGAGAACCA 550
184 hTrAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
551 CTGCTGGAGGTCTTAAACCTGCGACAGGAGAAAGAAATGAAAACTACACGA 600
201 SerLeuProArgAspThrSerAsnTyrSerAsnGlnPheGlnAlaGAspAs 217
601 AGCCTCCGCGGATACTAGTAAGTGTGCTTACCAATTTCAGAGAGACAA 650

```

```

217 nAlaArgSerSerLeuSerAlaSerHsPrometValGlyLysTrpLeug 234
  |||||
651 TGTGCTGCTGCTCTGAGTGCAGAGTCAACCCATGTTGGTGAAGTGGCG 700
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
701 AGAAACCAAGAACAGAGATGAGATGGAGACAGAAAGAGATTAACAGCGTGT 750
251 GluProValGlyHsAlaAspThrGlyLeuGluHsIleLeuProAsnPhse 267
751 GAACCTGTGGACATGCTGACACGGGTTTGGAGCATATACCAACTTTTC 800
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284
801 TCTGGATGATATGTAAGCTCTGAGAAATCCCAACGATGGAGGCGCTC 850
284 euGlyIleHsIleValValProPheSerAlaArgGlyLysArgThrLeuGly 300
851 TCGGAATCCATGTATGTCCTTTTCAGTCTCGAGGCGCGCAACCTGGGG 900
301 LeuLeuValLysArgLeuGlyLysGlyLysAlaGluHsIleGlnAsnLe 317
901 TTATATGTAACACGATGGAGAAAGTGTAAACCTGACATGAAATCT 950
317 uPheArgLysAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
951 TTTTCTGTAGATGATTTCTCATTTGTCAGAGATTATGATGGCGACCTTCA 1000
334 snArgArgPheGluGlnAlaGlnHsIlePheArgGlnAlaMetArgThr 350
1001 ATAGAAATTTTGAACACACCAACATATGTTCCCAAGCCATCCGTACA 1050
351 ProIleIleThrPheHsIleValValProAlaAlaAsnLysGluGlnArg 367
1051 CCCATCATTTGTTCCATGTTGCTCTCCACGCAATTAAGACAGATATA 1100
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384
1101 ACNACTATCCCAAGTGAAGAAATTAATCTATTCACCCGCTTTTACCC 1150
384 rAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHs 400
1151 CTGACAGCCACTATATTTGACACACAGAGTGTGACAGTGGCTTTCAC 1200
401 ThrValGlnArgAlaProArgLeuAsnHsProGluGlnIleAspSer 417
1201 ACGGTGAGAGAGACCCCGACTGAAACACCCGCTGAGCAGATAGACTC 1250
417 rHisSerArgLeuProHsIleSerAlaHsProSerGlyLysProProSerA 434
1251 TCACCTAAGACATCACTCATACGCAACCCCTCGGGAACCAACCATCCG 1300
434 lAProAlaSerAlaProGluAsnValPheSerThrThrValSerSerGly 450
1301 CTCGACCTCGGCACCTCAAGATGTAATTTAGTACGACTTAAGCAGTGGT 1350
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
1351 TATAACACCAAAAAATAGCAAGAGCTTAAATATCCAGCTTAAAGAAAG 1400
467 YThrGlnGlyLysGlyPheSerIleThrSerArgAspValThrIleGlyG 484
1401 TACAGAAAGTTTGGATTCACATCACTCTCCAAAGATGTAAACAATAGTGG 1450
484 lYserAlaProIleTyrValLysAsnIleLeuProArgLysAlaIle 500
1451 GCTCAGCTCCCAATCATGTGAAAAACATTTCTCCCGGGGCGGCCCATTT 1500
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517
1501 CAGATGGCGCACTTAAGCGAGAGACAGACTTATTAAGGTAAATGAGGT 1550
517 lAspLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSerT 534

```



```

|||||
1551 AGATTAGTGGGCAATCCCAAGAGAAAGTTGTCCTGTGGAAGCA 1600
534 hrlsMetGluGlyThrValSerLeuLeuValPheArgIngluAspAla 550
|||||
1601 CCAAGATGGAAGAACTGTGAGCTTCTGTCTTCCCGAGAAAGAGGCC 1650
551 PheHisProArgLuuLeuAsnAlaGluProSerGlnMetGlnIleProly 567
|||||
1651 TTCACCCCAAGGAACTGAAATGCAAGCCAAAGCAGATGCAATGCCAA 1700
567 sgluThrLysAlaGluAspGluAspIleValIleuThrProAspGlyThra 584
|||||
1701 AGAAGCAAGAGCAGAAAGATCAGATATGTTCTTACACCTGATGGCACCA 1750
584 rgluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
|||||
1751 GGAATTTTCGACATTTGAAATGCCACTTATATTCAGATCTCGAGGC 1800
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspIle 617
|||||
1801 CTGGTGTCAGTGTCAAAAGGTAAACGGTCAAAAGAACCAACCAAGATT 1850
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
|||||
1851 GGAATCTTTGTCAGTCCATTTATTAATGAGAGAGACGACATCTAAAGATG 1900
634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysLeu 650
|||||
1901 GAAGGCTTCGGGTGAATGATCAACTGATAGCACTAAATGCAATCCCTG 1950
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
|||||
1951 TTGGGCAAGCAAAACCAAGATGCCATGGAAACCTTAAGAAAGCTATGTC 2000
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684
|||||
2001 TACTGAAGCAATAAACGAGAAATGATCCAGCTTATTTGTGCAAGAGAA 2050
684 LeSerLysCysAsnGluLeuLysSerProGlySerProGlyProGlu 700
|||||
2051 TAAGCAAGTCOAATGAGCTGAATCACTGAGAGCCCTCGACCTGAG 2100
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
|||||
2101 CTGCCCATTTGAACACAGCTGTGATGATAGAGAACGAAGATTTCCCATTC 2150
717 rLeuArgSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
|||||
2151 CCTGTACAGTGGGATGAGGGCTGTGATGAAATCCCCACGCAAGAAATGCTG 2200
734 lAluSerArgIleMetGlyLysSerGlyLysArgGluLeuSerProThr 750
|||||
2201 CCTCACTGAGATAAATGGGTGATCACTGAATCACTGAGAGCCCTGCTACA 2250
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
|||||
2251 GTGAATATGCCCCCAAGATGACACTGTATATAGAAATGACAGAGTTGCC 2300
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784
|||||
2301 AGTGTCTCTCCACATCTCTGTACACAGTCTCTCCAGCTCCCAAGATG 2350
784 sPValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIle 800
|||||
2351 ATGTGGGGTTTGTGACGCGAGATGCTGTACTTGGGCCAAAGCTGCATC 2400
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
|||||
2401 AGTGATTCAGCCGACTGCTCTTGTAGTCCAGATGTGATCCACTGCTTGC 2450
817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
|||||
2451 TTTTCAACGAGAAAGATTGGAGCTCAGAGTATGTCAGAAAAACGACAA 2500
834 YsgInPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
|||||
2501 AGCAATTTTCAGATGCCGATCAATGGATTTGTTAAACACGAAATCA 2550
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
|||||
2551 AAAAGCATGTGATTTAGTATAGTGAAGACGACTAAACTCAATACAGTGA 2600
867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
|||||
2601 TGACCAAGAAAGCAGGTTCTCCACAGACAGATGTGGCTCTTCCCTGGGTC 2650
884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900
|||||
2651 TGAAAGAGTCAAGCTCGTTGGAGAGTCTGCAGACCGCAGTTGCCAGAGTG 2700
901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleArg 917
|||||
2701 ACTTGAATGGGATATTTCTTCCATGTCACAGGCGCGGATATATCAG 2750
917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyra 934
|||||
2751 AGGCAGGGGATGCATATGATGACAGCTTCAGAGCTGCATGCACAAATCTATG 2800
934 sPlysProAlaValAspAspAspAspGluGlyMetGluThrLeuGluGlu 950
|||||
2801 ATTAACCCGGGGTGTGATGATGATGATCAAGCATGGAGACCTTGAAAGAA 2850
951 AspThrGlnLysSerSerArgSerGlyArgGluSerValSerThrAlaSe 967
|||||
2851 GACACAGAAAGAAAGTTCAAGATCAGGAGAGAGCTGTATCCACAGGCAG 2900
967 rAspGlnProSerHisSerLeuGluArgGlnMetAsnGlyLysGlnLul 984
|||||
2901 TGATCAGCCCTTCCACCTCTGTGAGAGACAAATCAATGAAGCAACCAAGA 2950
984 YsgLysAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000
|||||
2951 AAGGTGATTAAGCTGATAGAAAGGATTAATGAAACCAAGAAAGGAAATGCT 3000
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
|||||
3001 AAAGATGAGATTAAGGAGAAAGATTAATGAAACCAAGAAAGGAAATGCT 3050
1017 uLysGlyLeuGlyLysMetPheArgPheArgLysHisArgLysAspL 1034
|||||
3051 GAAAGGCTTGGAGACATGTTTCAGGTTTGGCAAAATCGCAAAAGATGACA 3100
1034 YsIleGluLysThrGlyLysIleLysIleGlnLysSerPheThrSerGlu 1050
|||||
3101 AGATTGGAAGAACCGGTAAATATAAATACAGAAATCTTTACTTCGAA 3150
1051 GluGluArgIleArgMetLysGlnGluGlnGluArgIleGlnAlaLysTh 1067
|||||
3151 GAGAGAGATAGCAATGAAAGACAGAGACAGAGAGATTCOAAGCAAAAC 3200
1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyraIleGluIleG 1084
|||||
3201 TCGAGAAATTTAGGGAACGACAAAGCTCGAGAGCGGACTATGCTGAAATTC 3250
1084 lAspPheHisArgThrPheGlyCysAspAspGluLeuMetTyraGlyGly 1100
|||||
3251 AAGATTTTCTCGACACTTTGGCTGTATGATGATGATTAAGTATGGGGGA 3300
1101 ValSerSerTyrgLugLysSerMetAlaLeuAsnAlaArgProGlnSerPr 1117
|||||
3301 GTTCTCTTATGAAAGGTTCCATGGGCTCTCAACGCTAGACTCAGAGACC 3350
1117 oArgGluGlyHisMetMetAspAlaLeuTyraIleGlnValLysLysPro 1134
|||||
3351 ACGAGAAAGGCATATGATGATGCTTTGTATGCCCAAGTCAAGAAAGCCGC 3400

```

1134 rgAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150  
 3401 GGAATTCACAAACCTCAGCTGACAGAGTACAGATCACTCCAGCAAT 3450  
 1151 HisAspArg11eglnArgLeuArgLinglnPheglnGlnAlaLysLins 1167  
 3451 CAGATCGGATACAGCTGTGAGGCGAGAAATTCAGCAACCAACCAAGA 3500  
 1167 pGluAspVal1GluAspArgArgArgArgArgArgArgArgArgArg 3500  
 3501 TGAAGTGTAGAGATGCTCGGCGAGACTTAACTTTTGAGCAACCTGCG 3550  
 1184 roAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200  
 3551 CGAAGCGACGCGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1201  
 1201 ValGlnMetGlnArgGlnArgGlnGlnGlnGlnGlnGlnGlnGln 1217  
 3601 GTGCAGATGACGGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1217  
 1217 aglnArgGlnArgSerSerSerSerSerSerSerSerSerSerSer 3650  
 3651 CCAGCGCCAGTACAGCTCTCTGCTCGCAACAGCAAGCAAGCAAGCT 1234  
 1234 eValSerGlnAspSerTrpGlnGlnGlnGlnGlnGlnGlnGlnGln 3700  
 3701 CGGTCTCCAGAGCTCTGGGAGCGAGAGCTACTCCCTGGGAGAGCTTC 1250  
 1251 GlnSerAlaLysGlnAsnProArgArgArgArgArgArgArgArgArg 3750  
 3751 CAGAGTCCCAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1267  
 1267 nGlyTrpLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1284  
 3801 CGGCTACCTGGGAGGAGCATGGCTTCAAGCGCAGAGGCTATGCTGAA 3800  
 1284 lngLLeuLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3850  
 3851 AGGAGCTCTTCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1300  
 1301 GlnProProSerGlnGlyProSerAsnArgArgArgArgArgArgArg 3900  
 3901 CAGGCTCTTCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317  
 1317 nAspProSerTrpAlaProProLysGlyProPheArgGlnAspValPro 1334  
 3951 GGACCCCAAGTACGCCCCCTCCCAAGGGGCGCTTCCGGCAAGATGTC 1334  
 1334 roSerProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGlnLys 4000  
 4001 CCGCCCTCTCTCAGGTTGCGAGGCTGAAACAGACTTCAGACTCTGAGA 1350  
 1351 GlyArgProPheArgSer 1356  
 4051 GGGAGGCGCTCTATTC 4068

seq\_name: gb\_pr:AF467003

seq\_documentation\_block:  
 LOCUS AF467003 4062 bp mRNA linear PRI 07-FEB-2002  
 DEFINITION Homo sapiens partitioning-defective 3 protein splice variant b  
 ACCESION AF467003  
 VERSION AF467003  
 KEYWORDS AF467003.1 GI:18568345  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE  
 AUTHORS Gao, L., Macara, I. G. and Joberty, G.  
 TITLE Multiple splice variants of PAR3 and of a novel related gene,  
 PAR3i, produce functionally different proteins

REFERENCE 2 (bases 1 to 4062)  
 AUTHORS Gao, L., Macara, I. G. and Joberty, G.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-JAN-2002) Center for Cell Signaling, University of  
 Virginia, PO Box 800377, Charlottesville, VA 22908, USA

FEATURES  
 source  
 1..4062  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..4062  
 /gene="PAR3"  
 /note="PAR3"  
 1..4062  
 /gene="PAR3"  
 /note="PAR3"  
 /note="PAR3B; multi-PDZ domain containing protein;  
 /codon\_start=1  
 /product="partitioning-defective 3 protein splice variant  
 b"  
 /protein\_id="FAL76043.1"  
 /db\_xref="GI:18568345"  
 /translation="MKVYVCGRRVVPVPGDGHMKVFLIQAVTRKALAKDPNY  
 IFSELGRTNVVAFPOPYATSEIVTSVLRAMPHVRRSDPALIGLSTVSQSP  
 RSSEPRKMPSTTAFGLKONTAGSPKCDKRDENYKSLPRTSMNSNOFORDA  
 VNDGSPGLIIVVPSARGRTGLLVKRLRECGAAHEHNLFRENDCIVRINDGLN  
 REFEOAHMEKQAKRPIIWEHVPAKNEQVEQLSSEKNNYSRRSPDSQYIDR  
 SYNSAGLATVORALNHPPEQIDSHSRPHABSPGKPSAPASAPONVESTYVSG  
 YNTKIGRLNIDLRKTEGLSPSTSDVTGSLVAPLYVNIIPRGAIQDRLVLAG  
 DLLELVGVVLVGRSOEYVSLRSTKMGVSLVLPFOEAPFPRVETNAEBSQMP  
 KETAEDEDIVLPDGTRELPPEVPLNDSSAGIGVSGNGSKENHDLGIPYKSI  
 INGAASKIDRLRVNDOLIVANGESLICKTQMDAMETLRSMETENKMGMLQILVAR  
 RISKONELKSPSPGPPELPITETALDDERRISHSYSGIEGDESRRNALSLIR  
 KYOLSPVMPPODDPYITEDRLVLPPLHSDOSSSHDVFVTAADGTWAKAITS  
 DSADCSLSPDVPVATFQREGRGQSMSEKRTQFSDASQDLVTRKSKMDGLAD  
 ETKLNTVDDCKSSROVGPSLGLKSSLSLSLQRYAVETLNDIIPHPRPPIIR  
 ROMNGNOEKGDIDKDKTKCKEKRKDRKREKMKAKGMLKGLGDMFRGKRKRD  
 CIDELMTYGVSSYEGSMALNAROSPREGMADLYAQQKPRNSKSPVDSNRSTPS  
 NIDRTOQLRQEOEQAKODEVEDERRTYFEDPWNAPATQSRHSVSVYQKOR  
 OERSSQAOAROYSLPQSRKNAVSQDSWEDQNYSPGPGFOSAKENRISYSQGS  
 RNGLTGHEFNARVMELETOLRQDRRERDQMKKPPSBSQVSDYTKAVQDPSVAP  
 FRGPRQDVPSPSOVARLNRLQITPEKGRPYS"  
 814..1074  
 /gene="PAR3"  
 /note="Region: PDZ domain"  
 1378..1638  
 /gene="PAR3"  
 /note="Region: PDZ domain"  
 1765..2028  
 /gene="PAR3"  
 /note="Region: PDZ domain"  
 BASE COUNT 1218 a 922 c 1075 g 847 t  
 ORIGIN

misc\_feature  
 misc\_feature  
 misc\_feature

misc\_feature  
 misc\_feature  
 misc\_feature

alignment\_scores:  
 Quality: 1113.00  
 Ratio: 0.823  
 Percent Similarity: 99.779  
 Length: 1356  
 Gaps: 1  
 Percent Identity: 99.779

US-09-757-781-2 x AF467003  
 alignment\_block:  
 Align seg 1/1. to: AF467003 from: 1 to: 4062  
 1 MetLysValThrValCysPheGlyArgTrpArgValValProCysG1 17  
 1 ATGAAGTACCGTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 50

17 yAspGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34  
 |||||||  
 51 GGAGCGGCACATGAAAGTTTTCAGCCTCATCCAGAGCGGCGAACCCGCT 100  
 34 yTrArgLysAlaIleAlaLysAspProAsnYrTrpIleGlnValHisArg 50  
 |||||||  
 101 ACCGGAAGCGCATCGCCAAAGATCCAAACTAGTACAGTGCATCGC 150  
 51 LeuGlnHisGlyAspGlyLysIleLeuAspLeuAspAspIleLeuCysAs 67  
 |||||||  
 151 TTGGACATGGAGATGGAGGATACTAGACCTTGATGACATCTTTGGA 200  
 67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnGlnAsp 84  
 |||||||  
 201 TTGAGCAGACGATAAAGACACAGCTGTAGCAGTGTTCATGACGAGATC 250  
 84 roHisHisGlyLysAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100  
 |||||||  
 251 CACATCACGGAGGTGATGGCACACAGTCCAGTCCACGGTACCCAGAC 300  
 101 ProGlnIlePheGlySerGlnLeuGlyThrAsnAsnValSerAlaPheG 117  
 |||||||  
 301 CCAGAGATATTGGTAGAGCTTGACACCAACATGTCTCAGCCCTTCA 350  
 117 nProTyrGlnAlaThrSerGlnIleGlnValThrProSerValLeuArg 134  
 |||||||  
 351 GCCTTACCAACGACAAAGTGAATTTAGAGTCACACCTTCAGTCCGTGAG 400  
 134 LAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150  
 |||||||  
 401 CAAATATGCCCTTCATGTTCAGCAGATAGTGAACCCAGCTCTAATTGGC 450  
 151 LeuSerThrSerValSerAspSerAsnPheSerSerGlnGlnProSerAr 167  
 |||||||  
 451 CTCCTCCACTTGTGATGATGATGATTAATTTCTCTGAAGAGCCTTCAAG 500  
 167 gLysAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184  
 |||||||  
 501 GAAAAATCCCAACAGCTGTGTCAACACAGCTGGCTCTCAAGCAGAAC 550  
 184 hTrAlaGlySerProLysThrCysAspArgLysLysAspGlnAsnTyrArg 200  
 |||||||  
 551 CTGCTGGAGAGTCTTAAACCTGCACAGGAAGAAAGATGAAACATCACTA 600  
 201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217  
 |||||||  
 601 AGCCTCCCGGGGATACTAGTAAGTGTCTAACCAATTTCAAGAGACAA 650  
 217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234  
 |||||||  
 651 TGTGCGGCGCTCTGAGTGCACAGCAACCAATGTGGCGCAAGTGGCTGG 700  
 234 LulysGlnGlnGlnAspGlnAspGlyThrGlnGlnAspAsnSerArgVal 250  
 |||||||  
 701 AGAAACAGAAACAGATGAGATGGACAGAAAGAGATTAACAGTCTGTT 750  
 251 GlnProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267  
 |||||||  
 751 GAACCTGTGGACATGCTGACACAGGCTTTGGAGCATATACCCAACTTTTC 800  
 267 rLeuAspAspMetValLysLeuValGlnValProAsnAspGlyGlyProL 284  
 |||||||  
 801 TCTGATGATATGTGAAAGCTCTGAGAGTCCCAACGATGGAGGCGCTTC 850  
 284 euGlyIleHisValValProPheSerAlaArgGlyLysArgThrLeuGly 300  
 |||||||  
 851 TGGGAATTCACAGTAGTGCCTTTCAGTGTCTGAGGCGGACAAACCTGGGG 900  
 301 LeuLeuValLysArgLeuGlnLysGlyLysLysAlaGlnHisGlnLysAsp 317  
 |||||||  
 901 TTATTAGTAAACGATTGGAGAAAGGTGTAAAGCTGACATGAAATCT 950

317 uPheArgGlnAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334  
 |||||||  
 951 TTTTGTGAGAAATGATTCATGTGATGTCAGATTAATGATGGCAGCTTCGAA 1000  
 334 snArgArgPheGlnGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350  
 |||||||  
 1001 ATAGAAGATTGGTAACAGACAAACATATGTTTGGCCAAACCATGGGTACA 1050  
 351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGlnTyrG 367  
 |||||||  
 1051 CCCATCATTTTGGTTCATATGTGTTCTCTCGACGCAATTAAGACAGATACA 1100  
 367 uGlnLeuSerGlnSerGlnLysAsnAsnTyrTrpSerSerArgPheSerP 384  
 |||||||  
 1101 ACAACTATGCCCAAGTGAGAAACAAACATTAATTCACACCCGTTTATGCC 1150  
 384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlnLeuHis 400  
 |||||||  
 1151 CTGACAGCCAGTATATTGCAACAGAGGTGTGAACAGTGCAGGCGCTTCA 1200  
 401 ThrValGlnArgAlaProArgLeuAsnHisProProGlnGlnIleAspSe 417  
 |||||||  
 1201 ACGGTGCAGAGACACCCCGACTGAACACCCCGCTGACAGATAGACTTC 1250  
 417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434  
 |||||||  
 1251 TCACITCAAGACTACCTCATAGCGCACACCCCTCGGGAAAAACCACTCCG 1300  
 434 LProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450  
 |||||||  
 1301 CTCACAGCTCGGCACCTCGAATGTATTAGTACGACTGTAAGCATGTGT 1350  
 451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467  
 |||||||  
 1351 TATTAACACCAAAAAATAGGCAGAGAGCTTAATATCCACTTAAGAAAG 1400  
 467 yThrGlnGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484  
 |||||||  
 1401 TACAGAGGTTTGGGATTCAGCATCATCTCCAGAGATGTAAACAATAGTGG 1450  
 484 LysSerAlaProIleTyrValLysAsnIleLeuProArgGlnAlaIle 500  
 |||||||  
 1451 GCTCAGCTCCAAATCTATGTGAAAAACATTCCTCCCGGGGGGCGGCATT 1500  
 501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGlnValAsnGly 517  
 |||||||  
 1501 CAGGATGGCCGACTTAAGCAGAGACAGACTTAATAGAGGTAAATGAGAT 1550  
 517 LAspLeuValGlyLysSerGlnGlnGlnValValSerLeuLeuArgSerT 534  
 |||||||  
 1551 AGATTTAGTGGGCAAAATCCCAAGAGGAATGTGTTCCGCTTGAGAGACA 1600  
 534 hTrLysMetGlnGlyThrValSerLeuLeuValPheArgGlnGlnAspAla 550  
 |||||||  
 1601 CCAGATGGAAGAACTGTGAGCTTGTGCTTTCGACAGAGACGCGC 1650  
 551 PheHisProArgGlnLeuAsnAlaGlnProSerGlnMetGlnIleProLys 567  
 |||||||  
 1651 TTCCACCCCAAGGAACCTGAATGCAGAGCCCAAGCCAGATCAATTCGAAA 1700  
 567 sGlnThrLysAlaGlnAspGlnAspIleValLeuThrProAspGlyThrA 584  
 |||||||  
 1701 AGAAACGAAACAGAAAGATGAGATATTTGTTCTACACTGATGGACACA 1750  
 584 rGlnPheLeuThrPheGlnValProLeuAsnAspSerGlySerAlaGly 600  
 |||||||  
 1751 GGAATTTCTGACATTTGAAAGTCCCACTTAATGATTCAGAGATCTCAGGC 1800  
 601 LeuGlyValSerValLysGlnAsnArgSerLysGlnAsnHisAlaAspLe 617  
 |||||||  
 1801 CTTGGTGTACAGTCAAAAGGTAAACCGGTCAAAAGACAAACACAGCATTT 1850  
 617 uGlyIlePheValLysSerIleIleAsnGlyLysAlaAlaSerLysAspG 634

```

1851 GGGAAATCTTGTCAAGTCCATTATTAATGAGACGACATCTAAAGATG 1900
634 1YA9LleuArgValAsnAspGlnLeuIleAlaValAsnGlyLeuSerLeu 650
1901 GAAGGCTTCGGGTGATGATCAATCACTGATAGCAGTAATGCAATCCCTG 1950
651 LeuGlyPheThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
1951 TTGGGCAAGACAAACCAAGATGCAATGGAACCCCTAAAGAGTCTATGTC 2000
667 rThrGluGlyAsnLysArgGlyMetGlnLeuIleValAlaArgArgI 684
2001 TACTGAAGGCAATAAACGAGATATCTACACTATTGTTGCAAGAGGAA 2050
684 LeSerLysCysAsnGlnLeuLysSerProGlySerProGlyProGlu 700
2051 TAAGCAAGTGCAGATGAGCTGAAGTCACTGGAGAGCCCTGGACCTGAG 2100
701 LeuProIleGluThrAlaLeuAspAspArgLysArgIleSerHisSe 717
2101 CTGCCCATTTGAACACGCGTTGGATGATAGAGAACGAATTTCCCATTC 2150
717 rLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
2151 CCTCTACAGTGGGATGAGGCGCTTGATGAATCCCCAGACGAATGCTG 2200
734 lAlaSerArgIleMetGlyGluSerGlyLysTyrGlnLeuSerProThr 750
2201 CCTCCAGTGGGATTAATG.....GGTAATACCACTGCTCCCTACCA 2241
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767
2242 GTTAATATGCCCCAAGATGACCTCTCATTTAGAAAGATGACAGGTTC 2291
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 2291
2292 AGTGGTCTCTCACATCTCTCTGACACAGTCTCTTCAGCTCCCATGAG 2341
784 sPValGlyPheValThrAlaAspAlaGlyThrThrAlaLysAlaIle 800
2342 ATGGGGGTTTGTACGCGAGATGCTGTGTTGGCCCAAGGCTCCCAATC 2391
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
2392 AGTGAATCAGCCGACGCTCTTTGAGTCCAGATGTTGATCCAGTTCCT 2441
817 apHeGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
2442 TTTTCAACGAGAGATTTGGACGTCAAGATATGTCAGAAAAACGCCAA 2491
834 ySGLnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
2492 ACCAATTTTCAAGATGCCAGTCAATTGGATTTCGTTAAACACGAAATCA 2541
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
2542 AAAAGCATGATTTAGTATAGCTGACGAGACTAAACTCAATACAGTGA 2591
867 pasGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
2592 TGACCACAAAGCAGGTTCTCCGACGAGAGATGTGGTCTTCCTCCGGGTC 2641
884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900
2642 TGAAGAAAGTCAAGCTGTGTGAGACTGTCAGACCGCAGTGGCCGAGGTG 2691
901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleAr 917
2692 ACTTTGAAGGGGATATTCCTTCATCGTCCAGGCGCGGATTAATCAG 2741
917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerTyrA 934
|||||
2742 AGCGAGGGGATGCAATGAGACTTCAGAGCTGCCATTCGACAAATCTTATG 2791
934 sPlysProAlaValAlaAspAspAspGluGlyMetGluThrLeuGluGlu 950
2792 ATAAACCCGGGTAGATGATGATGATGAAGGATGAGACCTTGGAGAA 2841
951 AspThrGlnLeuSerSerArgSerGlyArgGluSerValSerThrAlaSe 967
2842 GACACAGAAAGAAAGTCAAGATCAGGAGAGAGTCTGATCCACAGGCGAG 2891
967 rAspGlnProSerHisSerLeuGluArgGlnMetAsnGlyAsnGlnGlu 2891
2892 TGATCAGCCCTTCCCACTCTGAGAGACCAATGAATGAACCAAGAGAG 2941
984 ySGLYAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000
2942 AAGGTATAGACTGATATGAAAAAAGATTAACCTGAAAAAGAAAGAG 2991
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
2992 AAAGTATAGATTAAGAGAGATTAATTAATTAAGCCAAAGAGGAAATGCT 1041
1017 uLysGlyLeuGlyAspPheArgPheGlyLysHisArgLysAspAspL 1034
3042 GAAGGCTTGGAGACATGTTCAAGTTCGTCGCAAAACATCGAAAGATGACA 1034
1034 ySILeGlyLysThrGlyLysIleLysIleGlnGluSerPheThrSerGlu 1050
3092 AGATTAGAAAGAGGATTAATTAATTAATTAATTAATTAATTAATTAAT 1050
1051 GluGluArgIleArgMetLysGlnGlnGluArgIleGlnAlaLysThr 1067
3142 GAGGAGAGGATACGAATGACAGACAGAGAGAGATTCACCAAAAC 3191
1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyrAlaGluIleG 3191
3192 TCGCAATATTAGGAGACGACAGCTCGACAGCGTCACTGCTGCAAAATTC 3241
1084 lAspPheHisArgThrPheGlyCysAspAspGluLeuMetLysGlyL 1100
3242 AAGATTTCATCGACATTTGGCTGTGATGATGATGATGATGATGATGATG 3291
1101 ValSerSerTyrGluGlySerMetAlaLeuAsnAlaArgProGlnSerP 1117
3292 GTTCTCTTATGAAAGTTCATGGCTCTCTCAACGCTAGACCTAGACACC 3341
1117 oArgGluGlyHisMetMetAspAlaLeuTyrAlaGlnValLysLysProA 1134
3342 ACCAGAAAGGCAATATGATGATGATGATGATGATGATGATGATGATGAT 3391
1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
3392 GGAATTTCCAAACCCCTACCTTACACAGTAAACAGATCAACTCTAGCANT 3441
1151 HisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAs 1167
3442 CATGATCGATACAGCGCTCTGAGGCAAGATTTCAACAAACCAACCAAGA 3491
1167 pGluAspValGluAspArgArgArgThrTyrSerPheGlnGlnInProTrp 1184
3492 TGAAGATGTGAAGATCGTCCGCGACACTATAGTTTGAACCAACCTGGGC 3541
1184 roAsnAlaArgProAlaThrArgLysSerGlyArgHisSerValSerValGlu 1200
3542 CGAAGCAGCGCGCGGCGAGCGCAGACGCGGCGACACTGGGTCTCGGTGAG 3591
1201 ValGlnMetGlnArgGlnArgGlnGluArgGlnLysSerGlnGlnAla 1217
3592 GTGCAATGCAACGCGCAGCGCAGAGAGAGCGAGAGCTCCACAGAGGC 3641
1217 agLnaArgGlnTyrSerSerLeuProArgGlnSerArgLysAsnAlaSer 1234
3642 CCAGGCGCAGTACAGCTCTCTCGCGCAAAACAGCAAAATATGCCAGCT 3691

```



```

134 IansMetProLeuHisValArgArgSerSerProAlaLeuIleGly 150
688 CAAATATGCCCTTCATGTCGACGCAAGTAGTGACCAGCTCAATTGGC 151
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
738 CTCCTCACTTCTGTCAAGTAGTAGTAATTTTCTCTGAAGACCTTCAG 167
167 GlyAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184
788 GAAAATCCACACAGCGGTGTCACACACACCTGGCTCTCAAGAGAACAA 837
184 hralaGlySerProLysThrCysAspArgLysLysAspGluAsnTrpArg 200
838 CTCCTGGAGTCTCTAAACCTCGACAGAGAAAGATGAAGAAACTACAGA 887
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217
888 AGCCTCCCGGGATACTAGTAACTGGTCTAACCAATTTCAGAGACAA 937
217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeuG 234
938 TCTGCTGCTGCTCTGAGTGCACCTCCCAATGTGGGCAAGTGGCTGG 987
234 IuLysGlnGluGlnAspGluAspGlyThrGluGlnAspAsnSerArgVal 250
988 AGAAACAAAGACAGATGAGGATGGACAGAAAGAGATTAACAGTCTGTT 1037
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
1038 GAACCTGTGGACATGCTGCACAGCGGTTTGGACCAATACCAACTTTC 1087
267 rIleuAspAspMetValLysLeuValGluValProAsnAspGlyGlyPro 284
1088 TCTGGATGATATGATGAAGCTCGTAGAAGTCCCAACAGATGAGGCGCTC 1137
284 euGlyIleHisValValProPheSerAlaArgGlyLysTrpThrLeuGly 300
1138 TGGGAATCCATGTAAGTCTTTCAGTCTGACGCGGCAAGACCTGGGG 1187
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGlnHisGluAsnLe 317
1188 TTATTATGAACGATTTGGAGAAAGTGTAAAGCTGAACATGAAAATCT 1237
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
1238 TTTTCGTGAGAAATGATTCATTTGTCAGATTAAATGATGGGACCTTCGA 1287
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
1288 ATGAGAAATTTGACACACACACATATGTTGCGCAAGCCATGCTTACA 1337
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnTrpG 367
1338 CCCATCATTTGTTCCATGTGTCCGCGACGCAAAATAAAGACATGATGA 1387
367 uGlnLeuSerGlnSerGluLysAsnAsnTrpTrpSerSerArgPheSerP 384
1388 ACAACTATCCCAAAGTGAAGAAACAAATTAATCAAGCCGTTTGAACC 1437
384 roAspSerGlnTrpIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
1438 CTGACAGCCGATATATGACACACAGAGATGACACATGCGAGGCTTCAC 1487
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
1488 ACGGTGACAGAGACACCCGACTACACCCCGCTGAGAGATAGACAC 1537
417 rHisSerArgLeuProHisSerSerAlaHisProSerGlyLysProProSe 434
1538 TCACCTCAAGACTACCTCATAGCGCACACCCCTGGGAAACCAACCATCG 1587

```

---

```

434 lAProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1588 CTCACAGCCTCGGACCTGCACCAATGATATTAGTACGACTGTAGCAGTCT 1637
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 1638
1638 TATTAACCCCAAAAAATAGCAAGAGGCTTAATATCCAGCTTAGAAGG 1687
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
1688 TACAGAAAGTTTGGATTTCAGCATCACTTCCAGAGATGTAACAATAGGT 1737
484 lYSerAlaProIleThrValLysAsnIleLeuProArgGlyAlaAlaIle 500
1738 GCTACACCTCCAAATCTATGTGAAAAACATCTCCCGGGGGCGGCATTT 1787
501 GluAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517
1788 CAGGATGGCCGACTTAAGCAGAGACAGACAGCTTAAGAGTAAATGGAGT 1837
517 lAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerT 534
1838 AGATTATGTGGCAAAATCCCAAGAGAAAGTTGTTCCGCTGTTGGAAGCA 1887
534 hrlYsMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 1888
1888 CCAAGATGGAAGAACTGTGAGCTTCGTGCTTTCGCGCAAGAAAGCGCC 1937
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 1938
1938 TTTCCACCCCAAGGAACTGAATGCAGACCAAGCAACATGCAGATTCCAA 1987
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThr 584
1988 AGAAACGAACACACAGATGAGATATTGTTCTTACACTGATGGCACCA 2037
584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
2038 GGGATTTCTGACATTTGAAGTCCCACTTAATGATTCAGATCTGCAGGC 2087
601 LeuGlyValSerValLysGlyLysAsnArgSerLysGluAsnHisAlaSp 617
2088 CTGGGTGTCAAGTCCAAAGGTACCGGTCAAAAGAAAGAACCAACCAAT 617
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAsp 634
2138 GGGANTCTTTGTCAGATCCATTATTAATGAGAGACACATCTAAAGATG 2187
634 lYArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerLeu 2188
2188 GAAAGCTTTCGGGTGAATGATCAACTGATAGCAGTAAATGGAATCCCTG 2237
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
2238 TTTGGCAAGACAAACCAAGATGCCATGGAACCCCTAAGACGTTATGTC 2287
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 2288
2288 TACTGAAGCAATTAACGAGATGATCCACCTTATTTGCAAGAGAGAA 2337
684 lSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 2338
2338 TAAAGCAAGTGCATAGCTGAACTCACCTGGAGCCCTCGAGACTGAG 2387
701 LeuProIleGluThrAlaLeuAspAspArgGluArgIleSerHisSe 717
2388 CTGCCATTTGAACACGCTTGATGATAGAAACGAAATTTCCCATTC 2437
717 rLeuThrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
2438 CTTCTACAGTGGATTTGAGGGCTTGATGAATCGCCACAGCAAAATGCTG 734
734 lLeuSerArgIleMetGlyLysSerGlyLysTrpGlnLeuSerProThr 750

```

2488 CCCCAGATGATATG.....GGTAAATACCACTGTCCTTCA 2528  
 751 ValasmetProGlnAspSerThrValIleIleGluAspAspArgLeuPr 767  
 2529 GTGATATGCCCAAGATGACACTGCTATTATGAGATGACAGGTTGCC 2578  
 767 OvalleuProHsiLeuSerAspGlnSerSerSerSerHsiAspa 784  
 2579 AGTGCTTCCTCCACATCTCTGACCAAGTCTCTTCCAGCTCCCATGATG 2628  
 784 spValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800  
 2629 ATGGGGGTTGTGACGGCAGATGCTGTTACTGGCCAGAGCTGCATTC 2678  
 801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValIleAl 817  
 2679 AGTATTCAGCCGACTGCTCTTGTAGTCCAGATGTTGATCCAGTTCTTGC 2728  
 817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834  
 2729 TTTTCAAGAGAGATTTGGACGTCAGAGTATGTCAGAAAAAGCCACAA 2778  
 834 ysgInPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850  
 2779 AGCAATTTTCAGATGCCAGTCAATTGATTTCTTAAACACGAAATCA 2828  
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867  
 2829 AAAAGCATGATTTAGTTATAGTGCAGACACTAAACTCAATACAGTGTGA 2878  
 867 PaSPglNLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884  
 2879 TGACACAGAAACAGAGTCTCCACAGAGATGTGGGTCCTCCCGGGTTC 2928  
 884 eulYsLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900  
 2929 TGAAGAAAGTCAAGCTCATTTGAGAGTGTGCAGACCGCAGTTGCCGAGGTG 2978  
 901 ThrLeuAsnGlyAspIleProPheHsiArgProArgProArgIleIleAr 917  
 2979 ACTTTGAATGGGGATATTCTTTTCATCTCCACAGCCCGCGATATATAG 3028  
 917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerTyra 934  
 3029 AGCGAGGGAGATGCAATGAGACTTCAGAGCTGCCATCGACAAACTGTATG 3078  
 934 spLysProAlaValAspAspAspAspGluGlyMetGluThrLeuGluL 950  
 3079 ATAAACCCGCGGTAGATGATGATGAAAGCATGGAGACCTTGGAGAA 3128  
 951 AspThrGlnGlnSerSerArgSerGlyArgGlnSerValSerThrAlaSe 967  
 3129 GACACAGAAAGAAAGTTCAGATCCAGGAGAGAGACTGTATCCACAGCCAG 3178  
 967 rAspGlnProSerHsiSerLeuGlnArgGlnMetAsnGlyAsnGlnGluL 984  
 3179 TGATCAGGCTTCCACCTCTCTGAGAGACAAATGAAATGAAACCAAGAGA 3228  
 984 ysgGlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000  
 3229 AAGGTGATTAACACTGATGAAAAAAGATMAACTGGAAAAAGAAAGAG 3278  
 1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017  
 3279 AAAGATGAGATTAAGAGAAAGATTAATGAAGCCAGAAAGGAAATCT 3328  
 1017 uLysGlyLeuGlnLysPhePheArgPheGlyLysHsiArgLysAspAl 1034  
 3329 GAAGGGCTTGGGAGACATGTTCAAGTTTGGCAAAACATCGAAAAAGATGACA 3378  
 1034 ysiIleGluLysThrGlyLysIleLysIleGlnGlnSerPheThrSerL 1050  
 |||||||||||||||||||||||||||||||||||||||||||||||||||

3379 AGATTGAGAAAAAGGTAATAATAAAAATACAGAAATCTTTACATCAGAA 3428  
 1051 GluGlnArgGlyLeuArgMetLysGlnGlnGlnGlnArgGlyIleGlnAlaLysTh 1067  
 3429 GAGGAGAGATTCGAATGACAGCAGCAGCAGCAGCAGATTTCAAGCCAAAC 3478  
 1067 rArgGluPheArgGlnArgGlnAlaArgGlnArgAspTyraLagIuIleG 1084  
 3479 TCGAAGATTTTGGGAAGACAGCAAGCTCGAAGAGCTGACTATGCTGAATTC 3528  
 1084 lAspPheHsiArgThrPheGlyCysAspAspGluLeuMetGlyrGlyL 1100  
 3529 AAGATTTTCATCGGACATTTGGCTGTGATGATGATTAATGATGAGGA 3578  
 1101 ValSerSerTyrgLugLysSerMetAlaLeuAsnAlaArgProGlnSerPr 1117  
 3579 GTTTCTTCTTATGAAGTTCCTCATGGCTCTCAACGCTACAGCTCAAGCCC 3628  
 1117 oArgGluGlyHsiMetMetAspAlaLeuTyraLagInValLysLysProA 1134  
 3629 ACGAGAAAGGCAATATGATGATGATGCTTTGTATGCCCAAGTCAAGAGCCGC 3678  
 1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150  
 3679 GGAATTCAAAACCTCACCTGTAGACAGTAAACAGATCAACTCTTAGCAT 3728  
 1151 HsiAspArgGlyIleGlnArgLeuArgGlnGlnGlnGlnAlaLysGlnAs 1167  
 3729 CATGATCGGATACAGCTCTGAGGCAAGATTTTCAGCAAGCAAGCAAGCA 3778  
 1167 pGluAspValGluAspArgArgArgThrTyrsSerPheGlnGlnProTrp 1184  
 3779 TGAAGATGAGAAAGATCGTCGGCGGACCTATAGTTTGAACAACCTGGC 3828  
 1184 rOAsnAlaArgProAlaThrGlnSerGlyArgHsiSerValGlu 1200  
 3829 CGAAGCAGCGCGGCGAGCAGCAGCGCGGCGACCTCGGTGTCCGTGGAG 3878  
 1201 ValGlnMetGlnArgGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 1217  
 3879 GTGCAGATGCAAGCGGACAGCGCAGCAGAGAGAGCGGAGAGCTCCACAGGC 3928  
 1217 aGlnArgGlnTyrsSerSerLeuProArgGlnSerArgLysAsnAlaSerS 1234  
 3929 CCAGCGCCAGTACACTCTCTCTCGGCAAGCAGAAATAATGCCAGCT 3978  
 1234 ervaSerGlnAspSerTrpGluGlnAsnTyrsSerProGlyLugLysPhe 1250  
 3979 CGGTCTCCAGACTCTTGGGAGCAGAACTACTCCCTGGGGAGAGCTTC 4028  
 1251 GlnSerAlaLysGluAsnProArgTyrsSerSerTyrgLingLysSerArgAs 1267  
 4029 CAGAGTCCCAAGAGAACCCCGAGTACTCCAGTTCACAAAGCTCCACAGAA 4078  
 1267 nGlyTyrsLeuGlyGlyHsiGlyPheAsnAlaArgValMetLeuGluThrG 1284  
 4079 CGGCTACTCGGAGAGACATGGCTTCAAGCCAGGGATCATGTGGAAGTTC 4128  
 1284 lngluLeuLeuArgGlnGlnGlnArgArgGlyGlnGlnGlnMetLysLys 1300  
 4129 AGGACTCTCTTCGCCAGGAACAGAGCGGAGACAGCAGCAGATGGAAGAAG 4178  
 1301 GlnProProSerGluGlyProSerAsnTyrsAspSerTyrsLysValGln 1317  
 4179 CAGCTCTCTTCGAGAGGGGCCAGCAACTATGACTGTATGAAGAGTCCA 4228  
 1317 nasPProSerTyraLapProProLysGlyProPheArgGlnAspValProp 1334  
 4229 GGACCCCAAGTTCAGCCCTCCCAAGGGGCCCTTCGGCAAGATGTGCCCC 4278  
 1334 roSerProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGluLys 1350  
 4279 CTTCCCTTCTCAGGTGCGAGGCTGAACAGACTTCAGACTCTCTGAGAA 4328  
 |||||||||||||||||||||||||||||||||||||||||||||||||||

```

1765..2028
/gene="PARD3"
/note="Region: PDZ domain"
BASE COUNT      1170 a      910 c      1048 g      832 t
ORIGIN

alignment_scores:
    Quality: 1025.00
    Ratio: 1.000
    Percent Similarity: 100.000
    Percent Identity: 100.000
alignment_start:
US-09-757-781-2 x AF467004
..
Align seg 1/1 to: AF467004 from: 1 to: 3960

1 MetlyValThrValCysPheglYArghrArValValProCysgl 17
1 ATGAAATGACCGCTGTGCTCGGACGGACCGGGGTGTGCTGGCGG 50
17 yAspglYHsmElyValPheSerLeuIleGlnGlnAlaValThrAgt 34
51 GGAAGCGCCACATGAAAGTTTTCACGCTTCATCCAGCAGGGGGTGACCCCT 100
34 yArGysAlaIleAlaLysAspProAsnYrTrpIleGlnValHisArg 50
101 ACCGGAGCGCATCGCCACAGAGATCCAAACTACGTGATACAGGTGCATCCG 150
51 LeuGlnHsGlyAspGlyGlyIleLeuAspLeuAspPilleuCyas 67
151 TTGGAACATGGAGATGGAGAAATACTAGACCTGTATGACATCTTGTGA 200
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluInasp 84
201 TGTAGCAGAGATGAAGAAAGACAGACTGTAGCTGTGTGATGACAGCAGATC 250
84 rHsHsLsGlyGlyAspGlyThrSerAlaSerSerTrpGlyThrGlnSer 100
251 CACATCAGCAGAGCTGATGCGACCGAGTCCAGTCCAGCGGTCCACAGACC 300
101 ProGluIlePheGlySerGlyLeuGlyThrAspAsnValSerAlaPheG 117
301 CCAGAGATTTTGGTAGTGAGCTTGGCACCAACAAATGTCACGCTTCA 350
117 nProGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
351 GCCTTACCAAGCAACAAGAAATTGAGTCAACCTTCACTCTTGCAG 400
134 lAsnMetProLeuHsValArgAspSerSerAspProAlaLeuIleGly 150
401 CAATATGAGCTCTTCATGTTCACGCCACTACTGTGCCAGCTCTCAATTGGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGlnProSerAr 167
451 CTCTCACCTTGTGAGTAGTAGTAATTTTCTCTGAGAGACCTTCAAG 500
167 gLysAsnProThrArgTrpSerThrThrAlaGlyPheLeuGlyGlnAsn 184
501 GAAATATCCACACGGCTGGTCAACAACACCTGCTCTCAAGCAGACACA 550
184 hAlaGlySerProLysThrCysAspArgLysValAspGlnAsnTyArg 200
551 CTGCTGGAGAGTCTTAAACCTCGACAGCAAGAAAGATGAAACTACAGA 600
201 SerLeuProArgAspHisSerAsnTrpSerAsnGlnPheGlnArgAsp 217
601 AGGCTCCGCGGGATACAGTACTGGCTTAAACCAATTTCCAGAGACACA 650
217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeuG 234
651 TGCTGCTGTCTCTGAGTGCACACTACCCCAATGATGGGCAAGTGGCTGG 700

```



234 IuysgIngluGlnAspGluAspGlyThrGluGlnAspAsnSerArgVal 250  
 701 AGAAAGAAAGACAGATGAGATGGACAGAAAGAGATTAACAGTCGTGT 750  
 251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267  
 751 GAACCTTTGGACATGCTGACACAGGGTTTGGAGCATATACCCAACTTTTC 800  
 267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyProl 284  
 801 TCTGATGATATGTGTAAGCTCTAGAAGTCCCAACGATGAGAGGCTC 850  
 284 euGlyIleHisValValProPheSerAlaArgGlyGlyArgThrLeuGly 300  
 851 TGGGAATCCATGTAGTCCCTTTCAGTCTCTGAGCGGCGAGAACCTTGCGG 900  
 301 LeuLeuValLysArgLeuGluLysGlyGlyLysAlaGlnHisGluAsnLe 317  
 901 TTATTAGTAAACAGATTGGAGAAAGTGGTAAAGCTGAACTGAATAAATCT 950  
 317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334  
 951 TTTTCGTGAGATATGATGTCATGTTCAGAGATTATGATGGGAGCTTGCA 1000  
 334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350  
 1001 ATAGAAATTTGAAACAGACACACATATGTTTCCCAAGCCATCGTACA 1050  
 351 ProIleIleThrPheHisValValProAlaAlaAsnLysGluGlnIleArg 367  
 1051 CCCATCATTTGGTTCATGTGTTCTCTGACCAATTAAGACAGCTATAGA 1100  
 367 uGlnLeuSerGlnSerGluLysAsnAsnIleTyrSerSerArgPheSerP 384  
 1101 ACACTATCCCAAGTGAGAAAGACAATTACTATTCAAGCCGTTTATGCC 1150  
 384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400  
 1151 CTGACACCGATATATTGACAAACAGAGTGTGAACAGTGCAGGCGCTTCAC 1200  
 401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417  
 1201 ACGGTGCAGAGACACCCCGACCTGAAACACCCCGCTGACGAGATAGACTC 1250  
 417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434  
 1251 TCACCTCAAGACTACCTCATAGCGCACACCCCTCGGGAACACACATCCG 1300  
 434 lAspAlaSerAlaProGlnAsnValPheSerThrValSerSerGly 450  
 1301 CTCGAGCCTCGGACCTCAGAAATGATTATTAGTACGACTGTAAAGCAGTGT 1350  
 451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysGly 467  
 1351 TATTAACACCAAAAATAATAGCAAGAGCTTAATATCCAGCTTAAGAAAGG 1400  
 467 yThrGluGlyLeuGlyPheSerIleThrSerArgPValThrIleGly 484  
 1401 TAGAAGAGTTTGGATTTCACATCAGCTTCCAGAGATGTAAACATAGTGT 1450  
 484 lysSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500  
 1451 GCTCAGCTCCCAATCTATGTGAAAAAATTTCCCGGGGGGGCGGCAATT 1500  
 501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVal 517  
 1501 CAGATGGCCCACTTAAGCGACGAGACGACTTATAGAGGTPAAATGAGGT 1550  
 517 lAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerT 534  
 1551 AGATTTATGTGGCAATCCCAAGAGAGAGTGTTCGCTGTGAGAAACA 1600  
 534 hrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550

1601 CCAAGATGGAAGGAAGCTGTAGCCTTCGTGCTTTCGCGCAGAAACACC 1650  
 551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567  
 1651 TTCCACCCCAAGGAACTGAATCGACAGCCAAAGCCGATGACGATTCCAA 1700  
 567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584  
 1701 AGAAACGAAACAGAAAGATGAGATATGTTCTTATACCTGATGGCACCA 1750  
 584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600  
 1751 GGAATTTCTGACATTTGAACTCCACCTTAATGATTCAGATCTCCACGC 1800  
 601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617  
 1801 CTTGGTGTCACTGTCAAGATGACCGGTCAAAAGAAACCAACGAGATT 1850  
 617 uGlyIlePheValLysSerIleIleAsnGlyValAlaAlaSerLysAspG 634  
 1851 GGGATCTTTGTCAAGTCCATTATTAATGAGAGACACACATTAAGATG 1900  
 634 lYArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650  
 1901 GAAGCTTCGGGTGATGATCAACTGATAGCAGTAAATGAGAAATCCCTG 1950  
 651 LeuGlyLysThrAsnGlnAspAlaMetGluThrIleuArgArgSerMetSe 667  
 1951 TTGGCAAGACAAACCAAGATGCCATGGAACCCCTTAAGAAAGTCTATGTC 2000  
 667 rThrGluGlnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684  
 2001 TACTGAAGCATTAACAGAGATATGATCAGCTTATGTTCGAAGAGAA 2050  
 684 lSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700  
 2051 TTAAGCAATGCAATGATGAGTCACTCGGAGGCCCGCTGACCTGAG 2100  
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717  
 2101 CTGCCCATTTGAACAGCCTTGATGATATGAGACACCAAAATTTCCCATTC 2150  
 717 rLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734  
 2151 CCTTACAGTGGGATTTAGGGGCTGTGATGAATCCCGACAGAAATGCTG 2200  
 734 lAluSerArgIleMetGlyGluSerGlyLysTyrGlnLeuSerProThr 750  
 2201 CCTTCAGTAGATATATGGGTGAGTCAGGTAAATATACACCTGCCCCAC 2250  
 751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767  
 2251 GTGAATATGCCCAAGATGACCTGTCAATTATAGAAATGACAGGTTGCC 2300  
 767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784  
 2301 AGTGTCTTCACATCTCTCTGACCAATCTCCTCCAGCTCCCATGATG 2350  
 784 sPValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIle 800  
 2351 ATGTGGGTTTGTGACGCGACATCTGTACTTGGGCCCAAGCTCCAAATC 2400  
 801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817  
 2401 AGTGATTCAGCCGACTGCTTTTGAATCCAGATGTTGATCCAGTTCCTGC 2450  
 817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGlyLysArgThrL 834  
 2451 TTTTCAAGAGAGAGATTTGGACGTCAAGATATGTCAGAAAAAGCACAA 2500  
 834 yGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850

```

2501 AGCAATTTTCAGATGCCAGTCAATTCGATTAAACAGCAAAATCA 2550
851 LysSerMetAspLeuGlyTlleAlaAspLThrLysLeuAsnThrValAs 867
2551 AAAAGCATGATTTAGGTAAGTACCTGACGAGACTAAACCAATACAGTGA 2600
867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
2601 TGACCAAGAAAGCAGGTCTCCACAGAGATGTGGTCTCCCTGGGCTC 2650
884 euLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGlyVal 900
2651 TGAGAGATCAAGCTCGTTGACAGAGCTCGACAGCGAGCTTCCGAGGTG 2700
901 ThrLeuAsnGlyAspLleProPheHisArgProArgProArgLleIleArg 917
2701 ACTTTGAAATGGGATATTCCTTCCATCGTCCAGCGCGGAGTAAATCAG 2750
917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaLleAspLysSerTyrA 934
2751 AGGAGGAGGATGCAATGAGAGCTTCAGAGCTGCCATCGCAAAATCTTAG 2800
934 sPlySProAlaValAlaAspAspAspGlyLysMetGlnThrLeuGlnGlu 950
2801 ATAAACCCCGGTAGATGATGATGAGAGCGATGAGACCTTGGAGAA 2850
951 AspThrGlnGlnSerSerArgSerGlyArgGlnSerValSerThrAlaSe 967
2851 GACACAGAGAAAGTTCAAGATCAGGAGAGAGTGTGTATCCACAGCCAG 2900
967 TAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlnGlnGlnGlu 984
2901 TGATCGAGCTTCCACTCTCTGGAGAGCAATGAAATGAAACCAAGAGA 2950
984 yGqlYAspLysThrAspArgLysLysAspLysThrGlyLysGlyLysLys 1000
2951 AAGGTATAGACTGATAGAAAAAGATAAACGAAAAAGAAAAAGAG 3000
1001 LysAspArgAspLysGlyLysAspLysMetLysAlaLysGlyMetLe 1017
3001 AAAGATAGAGATAGAGAGAGATAAATGAAGCCAGAAAGAGGAAATGCT 3050
1017 uLysGlyLeuGlnGlyAspMetPheArg 1025
3051 GAAGGCTTGGAGACATGTTCAAG 3075
seq_name: gb_pr:AF196186
seq_documentation_block:
LOCUS AF196186 3470 bp mRNA linear PRI 16-NOV-2001
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein
ACCESSION AF196186
VERSION AF196186.1 GI:13491611
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 3470)
AUTHORS Fang,C.M. and Xu,Y.H.
TITLE Down-regulated expression of atypical PKC-binding domain deleted
JOURNAL Cell Res. 11 (3), 223-229 (2001)
MEDLINE 21520080
PUBMED 11642408
REFERENCE
2 (bases 1 to 3470)
AUTHORS Fang,C. and Xu,Y.
TITLE Exon/Intron Structure and Splicing Variants of a Novel Human
JOURNAL Polarity Gene, hASP
REFERENCE 3 (bases 1 to 3470)
AUTHORS Fang,C. and Xu,Y.

```

```

TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) Laboratory of Molecular and Cellular
Oncology, Shanghai Institutes for Biological Sciences, Chinese
Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, P.R. China
FEATURES
source
1..3470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
288..3383
/contig="hasipsa"
/codon_start=1
/product="atypical PKC isotype-specific interacting
protein short variant"
/protein_id="AK27892.1"
/db_xref="GI:13491612"
/translation="MKTVVCFGRTRVVVPCGSGHMKVPSLIQAVTRRAIAKDPNY
MIOVRLBHDGGLDDLDLIDVADKDRIVAVFPDQDHHGSDTSASTQSPF
IFGSLGTNNVSARQPVQATSEIYVTSVLARNPDLVRRSSDPALIGLSTQSPF
SSEPSRKRPTRMSTTAGFLKONTAGSPKTCDRKDNYSLSLPRDSIMNSNOFORNA
VPNDGFLGIHVFPFSAGRTGLLVKRLKGGKAHEMLFRENDIIVINDGLARN
RREDOAHMFQAMRTPIIFHVVPANPKRDYDOLDSQEKNNYSSRFPSDSYTDNR
SVNSAGLTVQARPLNHPPEQIDSHSLRLPSHAPSGKPSAPASAPQNVFTVSSG
YNTKRIKRLNIOLKKEGLEGITRSQVYIGSAP IYKNIIPGAAIIOBRKLAG
DLTLEAVNDVLGKSOEEVSLRSTMEGTSLVFRQDAHPRELNAESOMQIP
KETRAEDEDIYLPDQGRREPLTEVPLNDGSGAGLVSVGNRSKRNHADLGIYVSI
INGGASRDRRLRNDDLIIVNGESLGTQNDAMETLRSKSTEGNKMAIOLIVAR
RISCHNELKSGSPPELPIETALDRERRISHSYSGIEGLDESPNNAAIRKMS
KYQSLPYNMPODDVTIIEDRLPLVPLPHISDSSSSSDGDEGFTVADGCAAAIS
DSADCSLSPDVPYLAFOREGFRQSRSEKRTKOFDSADOLPVTYRKSKMDGLAD
ETKLNITVDQKASPSRDVPSGLKSSLSLEIOTFAVAVYTLNGDIPHRPPIIR
GRGNESEFRAIRIISYDKPAVDDDCGMELEEDTDESSRSRESVSTASDQPSLE
R"
BASE COUNT 1003 a 811 c 934 g 721 t 1 others
ORIGIN
alignment_scores:
Quality: 781.00 Length: 1024
Ratio: 0.765 Gaps: 1
Percent Similarity: 99.707 Percent Identity: 99.707
alignment_block:
US-09-757-781-2 x AF196186
Align seg 1/1 to: AF196186 from: 1 to: 3470
1 MetLysValThrValCysPheGlyArgThrArgValValValProCysG 17
288 ATGAAGTGAACCGTGCTCGGAGGAGCCGGGTGGTCTGCGCGCGG 337
17 yAspGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34
338 GGAGGGCCACATGAAAGTTTTCAGCCATCCAGCAGCGGGTACCCTG 387
34 YArgGlyAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
388 ACCGGAAGGCCATCCCCAGAGATCCAACTACGTGATACAGGTGATCCC 437
51 LeuGlnHisGlyAspGlyLysLysLeuAspLeuAspAspIleLeuCysAs 67
438 TTGGAACATGAGAGATGAGAGAACTAGACCTTGATGACATTCCTTGG 487
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnGlnAsp 84
488 TGTAGCAGACGATTAAGACAGACTGTGATGATGATGATGATGATGAT 537
84 roHisHisGlyLysAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
538 CACATCAGCGAGGTGATGAGCAACCACTGCTCACGGGTACCCAGAGC 587

```

101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheGly 117  
 |||||  
 588 CCAAGATATTGGTAGTGGAGCTGGCACCAACAATGCTCAGCCTTCA 637  
 |||||  
 117 nProTyGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134  
 |||||  
 638 GCCTTACCAAGCAACAGTGAATGAGGTCAACCTTCAGTCCCTCGAG 687  
 |||||  
 134 IaasMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150  
 |||||  
 688 CAAATATAGCCTTTCATGTTGACGACGAGTATGACCCAGCTCTAATTGGC 737  
 |||||  
 151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167  
 |||||  
 738 CTCCTCACTCTGTGAGTAGATGAAATTTTCTCTGAAAGCCTTCAG 787  
 |||||  
 167 GlyAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184  
 |||||  
 788 GAAAAATCCCAACGCGTGTCAACACAGCTGCTCCCTCAAGCAGAAC 837  
 |||||  
 184 hTrAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyArg 200  
 |||||  
 838 CTGCTGGAGAGCTTAAAACTGCGACAGAAAGATGAAAACTACAGA 887  
 |||||  
 201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217  
 |||||  
 888 AGCCTCCCGCGGATACTAGTACTGCTCTAACCAATTCAGAGAACAA 937  
 |||||  
 217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeuG 234  
 |||||  
 938 TGTCTGCTGCTCTGTGAGTGCACCTCAACCATGTGGGCAAGTGGCTGG 987  
 |||||  
 234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250  
 |||||  
 988 AGAAACAAGAACAGATGAGGATGGACAGAAAGGATTAACAGTCGTCTT 1037  
 |||||  
 251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPheSe 267  
 |||||  
 1038 GAACCTGTGGACATGCTGACACGCGTTGGACATATACCAACTTTTC 1087  
 |||||  
 267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyGlyProL 284  
 |||||  
 1088 TCTGTGATGATATGTGTAAGCTCTGTAAGCTCCCAACGATGAGGGCTTC 1137  
 |||||  
 284 euGlyIleHisValValProPheSerAlaArgGlyLysArgThrLeuGly 300  
 |||||  
 1138 TGGGAATCCATGTAGTGCCTTTCAGTGTCCAGCGCGCAGAACCTGGGG 1187  
 |||||  
 301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisGluAsnLe 317  
 |||||  
 1188 TTATTAGTAAACGATTGGAGAAAGTGTGTAAGCTGMAACATGAAAAATCT 1237  
 |||||  
 317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyLysLeuArg 334  
 |||||  
 1238 TTTTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1287  
 |||||  
 334 snATGATGPhGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350  
 |||||  
 1288 ATAGAAGATTGTAACAGCAACACATATGTTCCGCAAGCCATCGGTACA 1337  
 |||||  
 351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnTyArg 367  
 |||||  
 1338 CCCATCATTTGTGTCATGTGTCTCCACCAATAAAGAGCAGTATAGA 1387  
 |||||  
 367 uGlnLeuSerGlnSerGluLysAsnAsnTyTrpSerSerArgPheSerP 384  
 |||||  
 1388 ACAACTATCCCAAGGTGAGAAACAATTAATTAATTAATTAATTAATTA 1437  
 |||||  
 384 roAspSerGlnTyrlLeuAspAsnArgSerValAsnSerAlaGlyLeuHis 400  
 |||||  
 1438 CTGACAGCCATATATTTGACAACAAGAGACTGTGAACAGTGCAGGCGTTGAC 1487

401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417  
 |||||  
 1488 ACGGTCGAGAGAGACACCCGACCTGAACACCCGCTGAGCAGATGAGACTC 1537  
 |||||  
 417 rHisSerArgLeuProHisSerSerAlaHisProSerGlyLysProProSera 434  
 |||||  
 1538 TCACCTCAAGACTTACCTCTACAGCGCACACCCCTGGGAAACACCATCCG 1587  
 |||||  
 434 IaProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450  
 |||||  
 1588 CTCACGCTCGGACACCTCAGAAATGATTTAGTACAGCTGTAAGAGTGTGT 1637  
 |||||  
 451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysGly 467  
 |||||  
 1638 TATAACACCAAAAAAATAGCAAGAGCCTTAATATACAGCTTAAGAAAG 1687  
 |||||  
 467 yThrGluGluLysLeuGlyPheSerIleThrSerArgPheValThrIleGly 484  
 |||||  
 1688 TACAGAAAGTTTGGATTCAGATCATCTTCCAGAGATGTATACAAATAGGTG 1737  
 |||||  
 484 LysSerAlaProIleTyArgValLysAsnIleLeuProArgGlyValAlaIle 500  
 |||||  
 1738 GCTCAGCTCCATCTATGTGAAAAACATTTCCCGCGGGGGCGGCATTT 1787  
 |||||  
 501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVal 517  
 |||||  
 1788 CAGGATGCGCCGACTTAAGCAGAGACAGACTTATAGAGGTAAATGAGAT 1837  
 |||||  
 517 LAspLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSerT 534  
 |||||  
 1838 AGATTATGTGGCAAAATCCCAAGAGAAAGTGTTCGCTGTGAGAACAA 1887  
 |||||  
 534 hTrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550  
 |||||  
 1888 CCAAGATGGAAGAACTGTGAGCCTTGTGCTTTCGCGCAGGAACACGCC 1937  
 |||||  
 551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLys 567  
 |||||  
 1938 TTCCACCCCAAGGAACCTGAATGCAGACCAAGCCAGATGCAGATTCCAA 1987  
 |||||  
 567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThr 584  
 |||||  
 1988 AGAAACGAAGACAGAAAGATGAGATATGTTGTTACACCTGATGGCACCA 2037  
 |||||  
 584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600  
 |||||  
 2038 GGGATTTCTGACATTTGAAGTCCACCTTAATGATTCAGAGATCTCAGGC 2087  
 |||||  
 601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617  
 |||||  
 2088 CTTGGTGTCAGTCTCAAGGTAAACCGGTCAAAAGAGAACACACGAGATTT 2137  
 |||||  
 617 uGlyIlePheValLysSerIleIleAsnGlyGlyValAlaSerLysAspG 634  
 |||||  
 2138 GGAATCTTTTGCAGTCCATTAATTAATGAGAGACAGCATCTTAAGAGTG 2187  
 |||||  
 634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysLeuSer 650  
 |||||  
 2188 GAAGCTTCGGGTGAATGATCACTGATAGCAGTAAATGGAATAATCCGTG 2237  
 |||||  
 651 LeuGlyLysTrpAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667  
 |||||  
 2238 TTGGCAAGACAAACCAAGATGCCATGGAACCCCTTAAGAAAGGTCTATGTC 2287  
 |||||  
 667 rThrGluGluLysAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684  
 |||||  
 2288 TACTGAAGCAATTAACAGAGATATCATCGCTTAATGTGTGCAAGAGAGAA 2337  
 |||||  
 684 LeSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700  
 |||||  
 2338 TAAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2387  
 |||||  
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717

```

|||||
2388 CTGCCCATGTAACACGCTTGAGATGAGAACGAGAAATTTCCCATTC 2437
717 TLeuTySerGlyIleGluIleuAspGluSerProSerArgAspAla 734
|||||
2438 CCTCTACAGTGGGATTCAGGGCTTGATGAAATGCCGACGAGAAATGCTG 2487
734 TAluSerArgIleMetGlyIleuSerGlyTyrcIleuSerProthr 750
|||||
2488 CCCCACTAGGATAATG.....GGTAATATACGAGCTGTCCTTACA 2528
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767
|||||
2529 GTGAATATGCCCAAGATGACACTGTCATTTATGAAGATGACAGGTGCC 2578
767 OValleuProProHisIleuSerAspGlnSerSerSerSerHisAspA 784
|||||
2579 AGTGGCTTCCCTCCACATCTCTGACAGTCCTTCCACCTCCACATGAG 2628
784 sPValGlyPheValThrAlaAspAlaGlyThrTPAlaAlaAlaIle 800
|||||
2629 ATTTGGGGTTTGTGACGGCAGATGCTGTACTTGGCCCAAGCTCAATC 2678
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValIleu 817
|||||
2679 AGTGAATTCAGCCGACTGCTTTGAGTCCAGATGTTGATCCAGTTCTGC 2728
817 aPheGlnArgGlyIleGlyPheGlyArgGlnSerMetSerGluTyArgTr 834
|||||
2729 TTTTTCACAGAGAGATTGAGCTGACGATGATGATGATGATGATGATG 2778
834 ySgIlnPheSerAspAlaSerGlnIleuAspPheValTyArgTySer 850
|||||
2779 ACCAATTTTCAGATGCAGTCAATTTGATTTCTTAAACGAGAAATCA 2828
851 LysSerMetAspLeuGlyIleAlaAspGluThrTyLeuAsnThrValAs 867
2829 AAAACCATGAGATTAGTATAGCTGACGACGAGCTAACTCATACAGTGA 2878
867 PAspGlnIleAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
2879 TGACCAGAAACAGGTTCTCCACGACAGATGAGGTCCTTCCCTGGGTC 2928
884 eLysIleTySerSerSerLeuGlnIleuSerLeuGlnIleuAlaIleAla 900
2929 TGAAGAACTCAAGCTCATTTGAGAGCTGACGACGCGAGTTGCCGAGTG 2978
901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIle 917
2979 ACTTTGAATGGGATATTTCTTTCCATGTCACGCGCGGATTAATCAG 3028
917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyra 934
3029 AGGCAAGGAGTGCATAGAGACTTCACAGCTGCCATCGCAAAATCTTATG 3078
934 sPLysProAlaValAspAspAspGluGlyMetGluThrLeuGluGlu 950
3079 ATTAACCCCGGATGATGATGATGATGATGATGATGATGATGATGATG 3128
951 AspThrGluIleuSerSerArgSerGlyArgGluSerValSerThrAla 967
3129 GACACAGAAAGTTCAGATGAGGAGAGAGAGTCTGTATCCACACCCAG 3178
967 rAspGlnProSerHisSerLeuGluArgIleMetAsnGlyAsnGlnGlu 984
3179 TGAATCAGCCCTCCACTCTCTGAGAGACAATGATGAGAAACCAAGACA 3228
984 ySgIleAspLysThrAspArgGlyLysAspLysThrGlyLysGluTylys 1000
3229 AAGGTATTAAGACTGATAGAAAAAGATMAAACTGAAAAAAGAAAG 3278
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysGlyMetLe 1017
|||||

```

```

3279 AAAGATACAGATTAAGAGAGATTAATGAAGCCAGAGAGGAAATGCT 3328
1017 uLysGlyLeuGlnIleuAspMetPhe 1024
|||||
3329 GAAGGCTTGGAGACATGCTTC 3350
seq_name: gb-pr:AK024668
seq_documentation_block:
LOCUS AK024668 2116 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21015 f15, clone CAE05730, highly similar to
ACCESSION AF252293 Homo sapiens PAR3 (PAR3) mRNA.
VERSION AK024668.1 GI:10437004
KEYWORDS oligo capping; f15 (full insert sequence).
SOURCE Homo sapiens primary endothelial cells of human coronary artery
cDNA to mRNA, clone_11b:CAE clone:CAE05730.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (siles)
REFERENCE
AUTHORS Kawabata,A., Hiki,J., T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2116)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
TITLE Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@elms.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan, cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
location/Qualifiers
1. 2116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="Primary endothelial cells of human coronary
artery"
/clone="CAE05730"
/clone_11b="CAE"
/note="cloning vector PME18SFL3"
1..2116
/note="highly similar to AF252293 Homo sapiens PAR3 (PAR3)
mRNA"
BASE COUNT 651 a 476 c 571 g 418 t
ORIGIN
alignment_scores:
Quality: 614.00 Length: 614
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x AK024668
Align seg 1/1 to: AK024668 from: 1 to: 2116
743 GlyLysTyrcIleuSerProThrValAsnMetProGlnAspAspThrVal 759
|||||
200 GGTAATATACGAGCTGTCCTTACAGTGAATATATGCCCCAAGATGACACTGT 249

```

759 IIEIIEGLIAspAspArgLeuProValLeuProHisLeuSerAspG 776  
 776 InsertSerSerSerHisAspAspValGlyPheValThrAlaAspAla 792  
 300 AGTCCTCTCCAGCTCCATGATGATGGGTTGTGACGGCAGATGCT 349  
 793 GLYThrTrpAlaLysAlaAlaIleSerAspSerAlaAspCysSerLeuSe 809  
 350 GGTACTTGGGCCAAGCTGCATGATGATTTCAGCCAGCTGCTTTGAG 399  
 809 rProAspValAspProValLeuAlaPheGlnArgGlyPheGlyArgG 826  
 400 TCCAGATGTGATCCAGTCTTCTGTTTCAACGAAAGGATTTTGAGCTC 449  
 826 InsertSerGlyLysArgThrLysGlnPheSerAspAlaSerGlnLeu 842  
 450 AGAGTATGTCAAGAAAGCACAAGCAATTTTCAGATGCCAGTCAATTG 499  
 843 AspPheValLysThrArgLysSerLysSerMetAspLeuGlyIleAlaAs 859  
 500 GATTGTGTTAAACAGCAAAATCAAAAGCATGATTTAGGTATAGCTGA 549  
 859 pGluThrLysLeuAsnThrValAspAspGlnLysAlaGlySerProSera 876  
 550 CGAGACTAAACTCAATCACTGATGACCGAAGAGGTTCTCCCGACA 599  
 876 rGAspValGlyProSerLeuGlyLeuLysLysSerSerSerLeuGlySer 892  
 600 GAGATGTGGTCCCTCCCTGGGTGTGAGAAAGTCAAGCTGTTGGAAGCT 649  
 893 LeuGlnThrAlaValAlaGluValThrLeuAsnGlyAspIleProPheH1 909  
 650 CTGACAGCCGACAGTGGCGAGGTGACTTGAATGGGATATCTCTTCCA 699  
 909 sArgProArgProArgGlyIleArgGlyArgGlyCysAsnGlySerPhea 926  
 700 TCGTCCAGCGCCGCGATTAATCAGAGCGAGGATGCAATGAGAGCTTCA 749  
 926 rGlnAlaAlaIleAspLysSerTyrAspLysProAlaValAspAspAsp 942  
 750 GAGCTGCCATCGACAAATCTTATGATTAACCCGGGTAGATGATGATGAT 799  
 943 GluGlyMetGluThrLeuGluGluLysPThrGluGluSerSerArgSerG1 959  
 800 GAAGGATGAGACCTTGGAAAGACACAGAAAGAAAGTTCAAGATCAGG 849  
 959 YArgGluSerValSerThrAlaSerAspGlnProSerHisSerLeuGluA 976  
 850 GAGAGAGTCTGTATCCACAGCCAGTGTACGCTTCCACTCTCTTGAGA 899  
 976 rGlnMetAsnGlyAsnGlnGluLysGlyLysPThrAspArgLysLys 992  
 900 GACAAATGAATGCAAACCAAGCAAGAAAGTATAGACTGATAGAAAGAG 949  
 993 AspLysThrGlyLysGlyLysLysLysAspArgAspLysGlyLysAspLys 1009  
 950 GATAAAACTGAAAAGAAAAGAAAGATAGAGATAGAGGAAGAGATAA 999  
 1009 sMetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArgP 1026  
 1000 AATGAAAGCCAAAGAAAGGAAATGCTGAAGGCTTGGGAGACATGTTCA 1049  
 1026 heGlyLysHisArgLysAspAspLysIleGlyLysThrGlyLysIleLys 1042  
 1050 TTGGCAAACTCCAAAAGATGACAGATTGAGAAAACGGTAAATATAAA 1099  
 1043 IleGlnGluSerPheThrSerGluGluGluArgIleArgMetLysGlnG1 1059  
 1100 ATACAGGAATCCTTTACATCAGAAAGAGAGAGATACGAATGAAGCAGGA 1149  
 1059 uGlnGluArgIleGlnAlaLysThrArgGluPheArgGluArgGlnAla 1076

1150 GCAGGAGAGCATTCACGCAAAACCTCGAGAAATTPAGGAGACAGCAAGCTC 1199  
 1076 rGluArgAspTrpAlaGluIleGlnAspPheHisArgThrPheGlyLys 1092  
 1200 GAGAGCGTACTATGCTGAATTCAGATTTTCAGATGGACATTTGGCTGT 1249  
 1093 AspAspGluLeuMetTyrGlyGlyValSerSerTyrGluGlySerMetAl 1109  
 1250 GATGATGATTAATGATGGGGAAGTTCTCTTATGAAAGTTCCATGCGC 1299  
 1109 aLeuAsnAlaArgProGlnSerProArgGlyLysHisMetLeuAlaL 1126  
 1300 TCTCAACCTAGACTCAGACCCACAGCAAGGGCATATGATGATGCTT 1349  
 1126 eUtyrAlaGlnValLysLysProArgAsnSerLysProSerProValAsp 1142  
 1350 TGTATGCCCAAGTCAAGAACCCGCGAATTCCAACCTCACCTGTACAC 1399  
 1143 SerAsnArgSerThrProSerAsnHisAspArgIleGlnArgLeuArgG1 1159  
 1400 AGTAACAGATCAACTCTAGCAATCATGATCGATACAGCGCTGAGCGA 1449  
 1159 nGluPheGlnAlaLysGlnAspGluAspValGluAspArgArgArgT 1176  
 1450 AGAATTTTCAGCAAGCAAAAGCAAGATGAAGATGTAGATGCTGCGCGA 1499  
 1176 hTyrSerPheGluGlnProTrpProAsnAlaArgProAlaThrGlnSer 1192  
 1500 CTTATAGTTTGGACMACCTTGCCGAGACGAGCGCCGCGACGAGAGC 1549  
 1193 GlyArgHisSerValSerValGluValGlnMetGlnArgGlnArgGlnG1 1209  
 1550 GGGGAGACACTGGGTGTCCTGGAGGTGAGATGACAGCGGCGAGCGCAGA 1599  
 1209 uGluArgGluSerSerGlnGlnAlaGlnArgGlnTyrSerSerLeuProA 1226  
 1600 GGAGCGCGAGAGCTCCACAGAGCGCCACGCGCAGTACAGCTCTGCGCTC 1649  
 1226 rGlnSerArgLysAsnAlaSerSerValSerGlnAspSerTrpGluGln 1242  
 1650 GGCAAAAGCAGAAAATGCCAGCTCGTCTCCAGAGACTTGTGGAGCAG 1699  
 1243 AsnTyrSerProGlyGluGlyPheGlnSerAlaLysGluAsnProArgTyr 1259  
 1700 AACTACTCCCTGGGGAAGGCTTCCAGAGTGCCTCAAGAAAGAACCCAGGTA 1749  
 1259 rSerSerTyrGlnGlySerArgAsnGlyTyrLeuGlyGlyHisGlyPhea 1276  
 1750 CTCAGACTACCAAGGCTCCAGGAACGGCTACCTGGAGAGACATGACTTCA 1799  
 1276 sNalArgValMetLeuGlnThrGlnGluLeuLeuArgGlnGlnArg 1292  
 1800 ACGCCAGGCTATGCTGGAACCTCAGAGCTCTTCCAGAGAAACAGAGG 1849  
 1293 ArgLysGluGlnGlnMetLysLysGlnProProSerGluGlyProSerAs 1309  
 1850 CGGAAGGAGCAGCAGATGAAGAGCAGCCCTCTCCAGGGGGCCACGACA 1899  
 1309 nTyrAspSerTyrLysLysValGlnAspProSerTyrAlaProPolyS 1326  
 1900 CTATGACTCGTATAAGAAAGTCCAGAGACCCAGTACCCCTCCCAAGG 1949  
 1326 LyrProPheArgGlnAspValProProSerProSerGlnValAlaArgLeu 1342  
 1950 GGCCCTTCCGCAAGATGTGCCCCCTCTCTTCCAGATTGCCAGGCTG 1999  
 1343 AsnArgLeuGlnThrProGluLysGlyArgProPheTyrSer 1356  
 2000 AACGACTTCAGACTCTGAGAAAGGAGGCCCTTCTATTC 2041

seq\_name: gb\_pr:AF332592

## seq documentation block:

LOCUS AF332592 2967 bp mRNA linear PRI 16-NOV-2001  
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein  
ACCESSION AF332592  
VERSION AF332592.1 GI:14579308  
KEYWORDS  
SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2967)  
AUTHORS Fang, C.M. and Xu, Y.H.  
TITLE Down-regulated expression of atypical PKC-binding domain deleted  
JOURNAL asip isoforms in human hepatocellular carcinomas  
MEDLINE Cell Res. 11 (3), 223-229 (2001)  
PUBMED 11642408  
REFERENCE 2 (bases 1 to 2967)  
AUTHORS Fang, C. and Xu, Y.  
TITLE Exon/Intron Structure And Splicing Variants of A Novel Human  
JOURNAL Polarity Gene, hasip  
REFERENCE 3 (bases 1 to 2967)  
AUTHORS Fang, C. and Xu, Y.  
TITLE Direct Submission  
JOURNAL Submitted (29-DEC-2000) Laboratory of Molecular and Cellular  
Oncology, Institute of Biochemistry and Cell Biology, Shanghai  
Institutes for Biological Sciences, Chinese Academy of Sciences,  
320 Yue-Yang Road, Shanghai 200031, P. R. China  
FEATURES  
source location/Qualifiers  
1..2967  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/map="10p11.2"  
1..2967  
/note="ASIPSB; alternatively spliced; similar to GenBank  
Accession Number AF196186"  
/codon\_start=1  
/product="atypical PKC isotype-specific interacting  
protein short variant b"  
/protein\_id="AAK69192.1"  
/db\_xref="GI:14579309"  
/translation="MKVTCFGRTRVYVPCGSHKVFSLIQAVTRKRIADKPY  
WIVNRLHEDGGILDDLDICDVADKDRLVAVDEDDPHHGGDTSASSTGOSPE  
IFSEELGNNVSAFOPYOATSEIYTPSVLANMPLHYRSDSDPALIGTSYDSNF  
SSEPRKNDPTFRSTAGFLKONTAGSPKCDRKDEYRGLPHDTNENNOQORDNA  
RSLSSAHMVKWLEKQEDQTEEDNSREVEPVGHGKAEHNFRENDCTIVINDQKIVE  
VPMDDGRLICIVHVPFSARGRTLGLVRLKKGKAEHNFRENDCTIVINDQKIVE  
SYNSAGLHVQAPRLNHPPIWHPVPAANKEDYDGLSEKNTYSSRFSDSOYIDNR  
YTKKIGKRLNIQKKTGEGEGFITSRDYTGSAATYVKNLIPKALIODRLKAG  
DRLIENGVDLVGKSOEYVSLRSTKEGTVALVPRQDAPHPELAEADDIVLT  
PDGREFLFEVPLNDSSAGTGYVGNKSKENHADLGFVSIINGCAASDGGTR  
VNDOLIAVNGESLIGKTNOMAEFLTKRSMTSEKNGMQLIVARISCNELKSPGS  
PPGELPIETALDPERIRISHTLSGIEGLDESPPNNAITRMGVYOSIPVNPQD  
DVIATEDRLPVLPHLSDSSSHDDVGFVADAGTAKKAIISNACSISPVNPQD  
VIAORREGFGQRLDETKLNTVDOKGSPSRGSLIKSSSLESLSIAVAEVL  
NGDIPRFRPRRIIRGRCNESFRALIDKSYKPAVDDEDEMETLEEDTESSRSR  
ESVYSTASDQSHSLEROMNGNOEKDTRKKDKTKGKKKKDKREKDKMAKKGMK  
GLDMEFLAKIKPKR"

BASE COUNT 890 a 653 c 776 g 648 t  
ORIGIN

## alignment\_scores:

Quality: 556.00 Length: 556  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment block:

US-09-757-781-2 x AF332592 ..

Align seq 1/1 to: AF332592 from: 1 to: 2967

```
1 MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17
|||||
1 ATGAAGTACCGCTGCTTCGAGCGAGCAGCCGGTGGTGGCCGCG 50
17 yAspGlyHisMetLysValPheSerLeuIleGlnAlaValThrArg 34
|||||
51 GGACGGCCCATGAAGAATTTCACGCTCATCCACAGCGGTCACCGCT 100
34 yTrArgValAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
|||||
101 ACCGGAGGCCATGCCAGAGATCCAAATCTAGTACAGGTCGTCGC 150
51 LeuGlnHisGlyAspGlyIleLeuAspLeuAspPheLeuCysAs 67
|||||
151 TTGGACATGGAGATGGAGCAATCTAGACCTTGATGACATTTGTGGA 200
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnHisArg 84
|||||
201 TGTAGCAGACGATTAAGACAGACTGTGACAGTGTATGACAGGATC 250
84 rHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
|||||
251 CACATCAGGAGGTATGGCACCATTCGACAGTCCACGCGTACCCAGAGC 300
101 ProGlnIlePheGlySerGlyLeuGlyThrAsnAsnValSerAlaPheG1 117
301 CCAGAGATTTTGTAGTACCTGGCACCACCAATCTCCACGCTTTCA 350
117 pTrpTrpGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
351 GCCTTACCAAGCAACAACTGAATGAGCCACCTCAGCTTCCTCGAG 400
134 IaAsMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
401 CAATATGCCCTTCATGTTGACGAGCAGTGTGACCCACCTCATNTTGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGlnGluProSer 167
451 CTCTCCACTTCTGATGATAGTATTTTCTCTGAGACCTTCCTCAAG 500
167 gLysAspProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsn 184
501 GAAAAATCCACACGCTGTTCAACAACAGCTGGCTTCTCAAGCAGAA 550
184 hTrAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyr 200
551 CTCTGGAGTCTTAAACCTGCGACAGAAAGATGAACCTACAGA 600
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217
601 ACCCTCCCGCGGATCTAGTACTGCTTAACCAATTTACAGAGACAA 650
217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeu 234
651 TGCTGCTGCTCTGTGAGTGCAGTCAACCAATGTTGGCAAGTGGCTG 700
234 LulysGlnGlnLysAspGlyAspGlyThrGlnGlnLysAsnSerArgVal 250
701 AGAAACAAGAAAGATGAGATGGACAGACAGAGATGAACAGTGTGTT 750
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
751 GAACCTTTGACATGCTGACACGGGTTTGGACATATACCAACTTTTC 800
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysPro 284
801 TCTGATGATATGTTAAAGCTCTGTAAGATCCCAACGATGAGGCTTC 850
284 euGlyIleHisValAlaProPheSerAlaArgGlyArgThrLeuGly 300
|||||
```

```

851 TGGGAATCCATGATGCTTTCAGTCTCGAGCGCGCAGAACCTCGGG 900
301 LeuLeuValIysArgLeuGluIysGlyIysAlaIleuHisGluAspLe 317
|||||
901 TTATATGAAAACGATTGGAGAAAGGTGGTAAAGCTGAACTGAAAAACT 950
|||||
317 uPheArgGluAsnAspCysIleValArgIleAsnAspCysIleAspLeuArg 334
|||||
951 TTTTCGTGAGAAATGATGATTCATGTCAGATTAATGATGCGACCTTCGAA 1000
|||||
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
|||||
1001 ATAAAGATTGGAACACACACACATATGTTTCCCAAGCCATCGCTTACA 1050
|||||
351 ProIleIleTrpPheHisValValProAlaIleAsnIysGluGlnIleArg 367
|||||
1051 CCCATCATTTGGTTCAGATGTGTTCTCTGACCAATAAAGACAGTATGA 1100
|||||
367 uGlnLeuSerGlnSerGluIysAsnAsnIleTyrTyrSerSerArgPheSer 384
|||||
1101 ACAACTATCCCAAGTAGAGAGAACAAATTAATTAACAAGCCGTTTACCC 1150
|||||
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
|||||
1151 CTGACAGCCAGTATATGACACAGAGAGTGTGAACAGTGCAGGCGCTTAC 1200
|||||
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
|||||
1201 AGCGTGCAGAGACGCCCGACTGACACCCGCTGAGCAGATGAGACTC 1250
|||||
417 HisSerArgLeuProHisSerAlaHisProSerGlyIysProProSerA 434
|||||
1251 TCACCTCAGAGTACCTCATGACGACACCCCTCGGGAACACACATCCG 1300
|||||
434 IapProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
|||||
1301 CTCAGGCTCGGACCTCAGATGATTTAGTACACTGTAAAGCAGTGTGT 1350
|||||
451 TyrAsnThrIysIysIleGlyIysArgLeuAsnIleGlnLeuIysAla 467
|||||
1351 TATTAACACCAAAAAATATGCAAGAGCGCTTAATATCCAGCTTAAGAAAG 1400
|||||
467 YThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
|||||
1401 TACGAGAGGTGGGATTCACATCATTCCAGAGATTAACAATAGGTG 1450-
|||||
484 IysSerAlaProIleTyrValIysAsnIleLeuProArgIleValAlaIle 500
|||||
1451 GCTCAGCTCCAACTATATGTAACAAATTCCTCCCGGGGGGCGCCACTT 1500
|||||
501 GlnAspGlyArgLeuIysAlaGlyAspArgLeuIleGlnIleValAsnGly 517
|||||
1501 CAGATGGCCGACTTAAGCAGAGACAGACTTATAGAGTAAATGAGAT 1550
|||||
517 IAspLeuValIleIysSerGlnIleGluValIleValSerLeuLeuArgSer 534
|||||
1551 AGATTATATGGCAAAATCCCAAGAGAGAGTGTTCCTGCTGTGAAACGA 1600
|||||
534 hTrpMetGluGlyThrValSerLeuLeuValPheArgGlnIleGluAspAla 550
|||||
1601 CCAAGATGGAAGAGACTGTGAGCCCTTGTGTCCTTCCGACGAGAAAGCC 1650
|||||
551 PheHisProArgGluLeu 556
|||||
1651 TTCACCCCAAGGGAAGCTG 1668
|||||
seq_name: gb_pr.AF332593
seq_documentation_block:
LOCUS AF332593 3822 bp mRNA linear PRI 16-NOV-2001
DEFINITION Homo sapiens atypical PKC isoform-specific interacting protein long
variant b mRNA, complete cds, alternatively spliced.
ACCESSION AF332593

```

```

VERSION AF332593.1 GI:14579310
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
REFERENCE
AUTHORS 1 (bases 1 to 3822)
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE Down-regulated expression of atypical PKC-binding domain deleted
asip isoforms in human hepatocellular carcinomas
JOURNAL Cell Res. 11 (3), 223-225 (2001)
MEDLINE 21520080
PUBMED 11642408
REFERENCE 2 (bases 1 to 3822)
AUTHORS Fang, C. and Xu, Y.
TITLE Exon/Intron Structure And Splicing Variants of A Novel Human
Polarity Gene, hasip
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 3822)
REFERENCE Fang, C. and Xu, Y.
TITLE Direct Submission
JOURNAL Submitted (29-DEC-2000) Laboratory of Molecular and Cellular
Oncology, Institute of Biochemistry and Cell Biology, Shanghai
Institutes for Biological Sciences, Chinese Academy of Sciences,
320 Yue-Yang Road, Shanghai 200031, P. R. China
FEATURES
source
1..3822
Location/Qualifiers
1..3822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
1..3822
/note="ASIPLB: alternatively spliced; similar to Genbank
Accession Number AF196185"
/codon_start=1
/product="atypical PKC isoform-specific interacting
protein long variant b"
/protein_id="AAK69193.1"
/db_xref="GI:14579311"
/translaton="MKVYVCFGRVYVPCGDDHMKVFSLIQOAVTRKAIKDPY
WIVHLEHEDGGILIDDLICDVAQDKDLVAFEDPDHGGDGTASCTGSPSE
IESGELGNNVSAFOPVQATSELEVTPSVLRANPLHYRSSDPALIGLSTVSDSF
SSSEPSKNTPTKMTSTTACGLKONTAGSPKTCDDKKENYSLIPRDSNWSNQQRDA
RSSLSASHPWVGWMLKEKODEDEGTEEDNSRVEPVGHADGLHIPFSLDDWKLVE
VPNDGPIGLIIVVPSARGRTGLLVKRLKGAKEHEMLFRENDCIVRIINDGLNR
RREFOAHGROARMPITWEHYVPAANKROYOLDSQEKNNYSSRFSPDSQIDNR
SVNSAGLHTYRORAPRLNHPPEOIDSRLPHSAHPSGKPPSAPASAPQNVFTVSSG
YNTKIKGRRLNIDLKGTGEGTSGTSRDTTIGSAPITYKNILPRCAIODEGLKAG
DRLLEVAGVDLVKRSQEEVSLRSTKMEGTVELVFRQDAPRPRELKADEDIYLT
PIDOTREFLEFEVPLNDGSAGLGVSVGNSKKNHADLGIFFVSIINGSAKSDGRAR
VNDQLAVNGESILGKTNDAMETLRKSMSTENKRMQILIVARISKNEKSPGS
PPPELPEITFALDDREERISLSYSGIEGDESPRSALSRIMGKTOLEPTVMPOD
DYVITIEDRLPYVPHLSQSSSSSHDDVEFTVADGTAKKAIISDADCSLSPNDP
VIAFORGEGROJADETKLNTVDDKAGSGSRDVGSLGKSSLSLESLDTAYAEVTL
NGDIPRPRPRRIIRGCMNESFRALIDKSYDPAVDDDEGMEITLEDDEESSRGR
ESVSTASIDQPSHLEROMNGNOSGDKTDDKKTGCKEKKDKDKDKKAKKAGMLK
GLDMDPFIQAKTEFEFERQARBERDYAEIODEHFTFGCDDDELAGVSSYGALNLR
POSPREGHMADALYAYQKPRNSKSPVDSNRSSTPSNHDRIOIRLBQFOOAKDEVE
DRRRYTFEQPWPNNAPARQSGRHSVVEVOMOROREEESGQAOAROVSSI,PROSR
KNASSYSDSEWENYSPGEGFOGAKENPRSSYOGSNGTGLGCHGARPMLETOELL
RQORRQEQQMKRQPSBEGPSNTDSIKVODPSYAPKGPFRDVPDPSPQVAKLNRL
QTPKGRPFS"
BASE COUNT 1114 a 885 c 1019 g 804 t
ORIGIN
alignment_scores:
Quality: 556.00 Length: 556
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x AF332593

```

Align seg 1/1 to: AF332593 from: 1 to: 3822

```

1 MetLysValThrValCysPheGlyValArgValValValProCysG1 17
1 ATGAAGTACCGTGTCTGCTGCGACGCGACCGGGGTGGTGGTCCGCGG 50
17 yAspGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34
51 GGACGGCCACATGAAAGTTTTCAGCTCATCCACAGCGGGGTGACCGCT 100
34 yTArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
101 ACCGGAAGCCATCCGCCAAGGATCCAAACTAGTGATACAGGTGCATGCG 150
51 LeuGluHisGlyAspGlyGlyIleLeuAspLeuAspPheIleLeuGlyCys 67
151 TTGGAAATGAGATGAGAGAAATCTAGACCTTGATGACATTTCTTTGTGA 200
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnAspP 84
201 TGTAGCAGACGATTAAGACACAGACTGGTAGCAGTGTGTGATGACAGATC 250
84 rHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
251 CACATCACGGAGGTGATGGCACCACTGCCAGTCCAGGGGTACCCAGAGC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
301 CCAGAGATATTGTGTAGTGGCTGGCACCAACTGTCTCAGCCCTTCA 350
117 nProGlyGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
351 GCGTTACACAGCAACAAAGTAATTGAGTGCACACTTCAGCCCTTCAAG 400
134 LaAsnMetProLeuHisValArgSerSerAspProAlaLeuIleGly 150
401 CAAATATGCTCTTCATGTTCACCCAGTAGGACCCAGCTCTAATTGGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerA 167
451 CTTCTCCACTTCTGTCAGTAGTAGTAATTTTTCCTGTAGAGAGCTTCA 500
167 gLysAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsn 184
501 GAAAAATCCACACGCTGTGTCAACACAGCTGCTCTCAAGCAGACA 550
184 hAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
551 CTGCTGGAGAGTCTAAACCTGCGACAGAAAGTAAGTAACCTACACA 600
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217
601 AGCTCCCGCGGATCTAGTAACTAGTGTCTAACCAATTTCAGAGACAA 650
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeu 234
651 TGTCTGCTGCTCTGTGAGTCCAGTCAACCAATGAGGCAAGTGGCTG 700
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
701 AGAAACACAGAACAGATGAGATGGACAGAAAGAGATTAACAGTGTGT 750
251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPhe 267
751 GAACCTGTTGACATGCTACACGGGTTTGGAGCATATATACCACTTTTC 800
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyGlyProL 284
801 TCTGGATGATATGTAAGCTGTGAAGTCCCAAGATGAGAGGCTTC 850
284 euGlyIleHisValValProPheSerAlaArgGlyGlyArgThrLeuGly 300
|||||

```

```

851 TGGGAATCCATGATAGTCCCTTTCAGTCCGAGCGCGAGAACCTTGGG 900
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGlnHisGluAsnLe 317
901 TTATTATGTAAGTAAAGATTGAGAAAGTGTAAAGTGAACATGAAATCT 950
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
951 TTTTGTGAGAAATGATTTCATTTGTGAGATTAATGATGGGACCTCGAA 1000
334 sNArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
1001 ATAGAAATTTTGAACACAGACACATATGTTTCCCAAGCCATGCTTACA 1050
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluIntArg1 367
1051 CCCATCATTTGTGTTCCATGTGTCTCGACGAAATTAAGACAGCATATGA 1100
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384
1101 ACAACTATCCCAAGTGAAGAAACAATTACTATTCAGCCGTTTATAGCC 1150
384 rAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
1151 CTGACAGCCAGTATATTGACACAGAGAGTGTGAACAGTGCAGGCTTAC 1200
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAsp 417
1201 ACGGTGAGAGACACCCGAGTACGAAACACCCGCTGACAGCATAGACTC 1250
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
1251 TCACCTAAGACATCTCATATGACACACCCCTCGGGAATACCATTCG 1300
434 LaProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1301 CTCCACCTCCGCGACCTCAGATGATATTAGACACTGTAAACAGATGT 1350
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysG1 467
1351 TATTAACCCAAAATAATAGCAAGAGGCTTAATATCCAGCTTAAGAAAG 1400
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
1401 TACACAAAGTTGGGATTCACATCAGCTTCACAGATGTAACATAGGTG 1450
484 ySerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
1451 GCTCAGCTCCATCTATGTGAAAAACATTCCTCCCGGGGGCGGCATT 1500
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGly 517
1501 CAGGATGGCGACTTAAGCAGAGACAGACTTAATAGAGTAATGAGAT 1550
517 LaSpLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerT 534
1551 AGATTTAGTGGGCAAAATCCCAAGAGAAAGTTTGTGCTGTTGAGAGACA 1600
534 hTyrMetGluGlyThrValSerLeuValPheArgGlnGluAspAla 550
1601 CCAAGATGGAAGAACTGTAGAGCTTCTGTGCTTTCGCCAGGAAGACGCC 1650
551 PheHisProArgGluLeu 566
1651 TTCACCCCAAGGGAAGCTG 1668

```

seq.name: qb\_pr:AF467006

seq.documentation\_block:

LOCUS AF467006

DEFINITION Homo sapiens partitioning-defective 3 protein splice variant f (PARD3) mRNA, complete cds, alternatively spliced.

ACCESSION AF467006

3933 bp

mRNA

linear

PRI 07-FEB-2002



VERSION AF467006.1 GI:18568351  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE  
 1 (bases 1 to 3933)  
 Gao, L., Macara, I.G. and Joberty, G.  
 Multiple splice variants of PAR3 and of a novel related gene,  
 PAR3L, produce functionally different proteins  
 JOURNAL  
 REFERENCE  
 2 (bases 1 to 3933)  
 Gao, L., Macara, I.G. and Joberty, G.  
 Direct Submission  
 Submitted (10-JAN-2002) Center for Cell Signaling, University of  
 Virginia, PO Box 800577, Charlottesville, VA 22908, USA  
 FEATURES  
 source  
 1..3933  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 gene  
 1..3933  
 /gene="PAR3"  
 /note="PAR3"  
 CDS  
 1..3933  
 /gene="PAR3"  
 /note="PAR3f: multi-PDZ domain containing protein;  
 alternatively spliced"  
 /codon\_start=1  
 /product="partitioning-defective 3 protein splice variant  
 f"  
 /protein\_id="AF467006.1"  
 /db\_xref="GI:18568352"  
 /translation="MKVTCFGKTRRVVPCGDSHMKVFSLIQQAVTRYRAIKADPNY  
 WIOVHLEHGDGIDLDLDDICDVADDDKDLVAVPEODPHGGDGTASSTQSE  
 ITGSELTNNVSAFOPYQATSEIEVPSVLRAVMPVLRSSDPALIGSTSDSN  
 SSEPSRKRPTRKSTAGFLKQNTASPTCDKREKENTRSLEPRDSNNMSNOQRDA  
 RSLSSHPVGVGMEKQODEDETEDSRVEPVHADGLSHINFSLDNKLVE  
 VPNDGGPILIHVVPFSAKGRITGLLVKLEKGAKEHLENFENDCIYVINGDLN  
 RREFQAHMFRQAMRPIIMFHVVPANKEQYQLQSEKNYSSRPSDSQYIDNR  
 SVNSAGLHVORAPRLNHPPEQIDSHSLPSHSAHPGKPSAPASAPQNVETTSVG  
 YVTKKIGKRLNIOLEKKGTEGLGFSITSDVTIGGSAPIYKNTLPPGAALODRLAK  
 DRLIENVGVDAVKGSOEYVSLSTRKMGTVSLVIFROEDAHPRELAEDDIYLT  
 PGTRFELTFEVPVLDNDSAGLVSYGNRSKNADLGLFVASTIINGNAGSGLR  
 VNDQLAVGESLGLKTNODAMETLRSMSTBENKGMQLIVARI SKCNELKSQS  
 PPGPELPIETALDREKRISHSLYSGIEGLDESPNNALSRIMGRYQSLVPMFD  
 DVIILEDRLPYLPALSDOSSSSHDVGFYADAGTAKAKAISDAQCSLSDVDP  
 VLAFOREGFGROIADETKLTNTVDOKAGSPSRVPSGLIKSSLSLEOTAAVETL  
 NGDIPHRPRPTIRIGRCNESPRAIDKSYDPRVADDDDEGKETLEEDTESSRGR  
 EYSTASDPSHSLERQMGNGNKGKTDKDKTKERKKDKEDKKAKKMLK  
 GLGDMERFGKHRRDKLERTGKIKIQSEFTSEERI RMQOERIQAKTREFEQRAR  
 ERDYAEIIDFHRTFGCDDELMTGVSYSKSMALNARPOS PRGHMMADALYAQVKPR  
 NSKSPVDNRSSTPNHRIQRLROFOQAQODEVDRETRTSFQPMWNAAPGOS  
 GHSVSEVEOMOROREERESSQAOQROFQSSLPORAKNASISQOSWRENYDGGF  
 OSAKEPRYSSTYOGSRNGYLGGHGFNARVMELEQLROEQRKEDQMKQPSSEGS  
 NVDSTIKVDDPSTAPKPPRDVPPSPSOVAVNLRLQTEKRPRTYS"  
 814..1074  
 /gene="PAR3"  
 /note="Region: PDZ domain"  
 1378..1638  
 /gene="PAR3"  
 /note="Region: PDZ domain"  
 1726..1989  
 /gene="PAR3"  
 /note="Region: PDZ domain"  
 BASE COUNT 1164 a 899 c 1049 g 821 t  
 ORIGIN  
 alignment\_scores: 556.00 Length: 556  
 Quality: 1.000 Gaps: 0  
 Ratio: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-757-781-2 x AF467006 ..  
 Align seg 1/1 to: AF467006 from: 1 to: 3933  
 1 MetLysValThrValCysPheGlyArgThrArgValValProCysGI 17  
 1 ATGAAAGTGACCGTGTGCTTCGGACGAGCCCGGTGCTGCTCCGTGCGG 50  
 17 yAspGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34  
 51 GGACGGCCACATGAAAGTTTCAGCCTCATCCAGAGGGGGTGACCGCT 100  
 34 yTArgLysAlaIleAlaLysAspProAsnYTrpIleGlnValHisArg 50  
 101 ACCGGAAGCCATCCGCAAGATCCAACTACTGATGTACAGGTGATGCC 150  
 51 LeuGluHisGlyAspGlyGlyIleLeuAspLeuAspAspIleLeuCysAs 67  
 151 TTGGAACATGGAGATGAGGAATACTAGACCTTGATGACATTCCTTTGTA 200  
 67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84  
 201 TGTACAGACGATAAAGACAGACTGTAGTACAGTGTGTGAGCAGGATC 250  
 84 rOHSHisGlyLysAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100  
 251 CACATCAGCAGGATGATGACACAGTGCAGTCCAGCGGTACCCAGAC 300  
 101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheGI 117  
 301 CCAGAAATATTGTTGATGAGTGTGGCACCACCAATGTTCTCAGCCTTCA 350  
 117 nProGlyGlnAlaThrSerGluIleGluValThrProSerValIleuArg 134  
 351 GCCTTACCAAGACACAGGAAATGAGTCAACCTTCATGCTTCGCGAG 400  
 134 IaAsnMetProLeuHisValAlaArgSerSerAspProAlaLeuIleGly 150  
 401 CAATATGCTCTTCATGTTGACAGCAGTACAGCCACTTAATTTGTC 450  
 151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167  
 451 CTCTCCACTTCTGTCAGTATGATATTTCTCTGTAAGACCTTCAAG 500  
 167 gLysAsnProThrArgTrpSerThrAlaGlyPheLeuLysGlnAsnT 184  
 501 GAAATATCCACAGCTGTCTCAACACAGCTGGCTTCTCAAGCAGAAC 550  
 184 hTAlaGlySerProLysThrCysAspArgLysLysAspGluAsnYTrArg 200  
 551 CTGCTGGGAGCTCTAAACCTCGACAGGAAAGATGAAACATACAC 600  
 201 SerLeuProAlaArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217  
 601 AGCCTCCCGCGGATACTAGTAACTAGTGTACCAATTTCAAGAGACAA 650  
 217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234  
 651 TGTCTGCTGCTCTGTGAGTCCAGTCCAAATGGGTGGGCAAGTGGCTGG 700  
 234 LuLysGlnGluGlnAspGluAspGlyThrGluGluLysAspAsnSerArgVal 250  
 701 AGAAACAGAAACAGATGAGATGGAGACAGAAAGATATACCACTGCTGT 750  
 251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPheSe 267  
 751 GAACCTGTTGGACATGCTGACACAGGTTTGGAGCATATACCAACTTTTC 800  
 267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284  
 801 TCTGATGATATGTTAAGCTCTGTAGAAAGTCCCAACGATGAGAGGCTTC 850

284 enuGlylleHISValValProPhSeSerAlaArgGlyIArgThrLeuGly 300  
 851 TGGGAATCCATGTAGTCTTTCAGTCTCGAGGCGGACACCTGCGG 900  
 301 LeuLeuValLysArgLeuGlyLysGlyGlyLysAlaGlnHisGlnLeu 317  
 901 TTATTTAGTAAACGATGTGAGAAAGGTGTAAAGCTGAACATGAAATCT 950  
 317 uPheAArgGlnAspCysIleValArgIleAsnAspGlyAspLeuArg 334  
 951 TTTTCGTGAAATGATGTCATGTCAGAGATTAATGATGAGGACCTTGAA 1000  
 334 snArgArgPheGlnGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350  
 1001 ATGAGAGATTTGAACAAGCACACATATGTTTCGCCAAGCATGCGTACA 1050  
 351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGlnGlnArg 367  
 1051 CCCATCATTTGGTCCATGTGGTTCCTGCGACAAATTAAGACAGCATATGA 1100  
 367 uGlnLeuSerGlnSerGlnLysAsnAsnLysArgSerSerArgPheSer 384  
 1101 ACAAGTATCCCAAGTGAAGAAACAATTAATCAAGCCGTTTATGCC 1150  
 384 roAspSerGlnLysIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400  
 1151 CTGACAGCCAGTATATTGACMACAGAGTGTGAACAGTGCAGGCGCTTAC 1200  
 401 ThrValGlnArgAlaProArgLeuAsnHisProProGlnGlnIleAsp 417  
 1201 ACCGTGACAGAGACACCCCGACTGACACCCCGCTGAGCAGATAGACTC 1250  
 417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSer 434  
 1251 TCACTACAGACTACCTCATAGCGACACCCCTCGGGAACCAACATCCG 1300  
 434 lArProAlaSerAlaProGlnAsnValPheSerThrValSerSerGly 450  
 1301 CTCACAGCTCGGACCTCAGAAATGATTATGTAGTACGACTGTAAACAGTGT 1350  
 451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467  
 1351 TATTAACACCAAAAAATAGSCAAGAGCTTAATATCCAGCTTAAGAAAGC 1400  
 467 YThrGlnGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484  
 1401 TACAGAAAGTTTGGATTCAGATCACTTCCAGAGATGTAAACAATAGTGT 1450  
 484 lYserIleAlaProIleLysValLysAsnIleLeuProArgGlyAlaIle 500  
 1451 GCTCAGCTCCAAATCTATGTGAAAAAATCTTCCCGCGGGCGGCGCATT 1500  
 501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGlnValAsnGlyVal 517  
 1501 CAGAGTGGCGCACTTAAGCGAGAGACAGACTATACAGGTAATGAGAT 1550  
 517 lAspLeuValGlyLysSerGlnGlnGlnValValSerLeuLeuArgSer 534  
 1551 AGATTAGTGGCAAAATCCCAAGAGAAAGTTGTTCCGCTGTGAGAAACA 1600  
 534 hrLysMetGlnGlyThrValSerLeuLeuValPheArgGlnGlnValAsp 550  
 1601 CCAAGATGGAAGAACTGTAGCCCTTCTGCTTTCGCGAGAAAGACGCC 1650  
 551 PheHisProArgGlnLeu 556  
 1651 TTCACCCCAAGGAGACTG 1668

seq\_name: gb\_pr:AF252293

seq\_documentation\_block:

LOCUS AF252293 3801 bp mRNA linear PRI 10-JAN-2002  
 DEFINITION Homo sapiens partitioning-defective 3 splice variant c (PAR3) mRNA,

complete cds, alternatively spliced.  
 AF252293  
 VERSION AF252293.1 GI:8037914

SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 3801)  
 Jobery,G., Petersen,C., Gao,L. and Macara,I.G.  
 The cell-polarity protein Par6 links Par3 and atypical protein

JOURNAL Nat. Cell Biol. 2 (8), 531-539 (2000)  
 MEDLINE 20394296  
 PUBMED 10934474

REFERENCE 2 (bases 1 to 3801)

AUTHORS Gao,L., Macara,I.G. and Jobery,G.  
 TITLE Multiple splice variants of Par3 and of a novel related gene,  
 Par3L, produce functionally different proteins

JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 3801)  
 Jobery,G.M., Petersen,C.D., Gao,L. and Macara,I.G.  
 TITLE Direct Submission

JOURNAL Submitted (05-APR-2000) Markey Center for Cell Signaling,  
 University of Virginia, West Complex, P.O. Box 800577,  
 Charlottesville, VA 22908-0577, USA

FEATURES  
 source 1..3801  
 Location/Qualifiers

1..3801  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="kidney"  
 1..3801  
 /gene="PAR3"  
 1..3801  
 /gene="PAR3"  
 /note="multi-PDZ containing protein; PAR3; alternatively  
 spliced"

/codon\_start=1  
 /product="partitioning-defective 3 splice variant c"  
 /protein\_id="AF71530.1"  
 /db\_xref="GI:8037915"

/translation="MKVVCGRPRPVVPCGDGHHKPSLQOATRYRKAKDPNY  
 WQVRLRHGDCGILDDLDLDVADDKDRLVAVDEDDPHHGCGTSASSTGSPSE  
 IFGSELGNNVSAFQPYQATSEITEVPSVLNANMLHRRSSDPALIGLSTVSQSNF  
 SSEPSRKNPTKMSSTAGFLKONTAGSPCTDRKDEDEEDNSVEEYVGHADTLEH  
 IPNEFLDMVRLKEVYNDGPGELGIVHVEFSARGRTDLVLRLEKGAHEHNTFRE  
 NDCIVRINDGDLNRFRFOAOHMFEOAKRTPIIMHVPAANKREQEDLOSSEKNY  
 SSRSPDSQOYIDNREVSNAAGLHTVORARPLNHPROIDSHSLPSAHPGKPPAPA  
 SAPQVFEFTTSSCTNTKICKRLNIQAKTGEGCFSTSDTYITGSAPIYVNNIL  
 PRGAIIQGRKLAKGRLIEVNGVDLVGSOEYVSLNSTKMEGVSLVLRQEDAFH  
 PRELAEDEDIVLTPDGTREFLTFEVPVPLNDSSAGLVSKGNSRKENHADGIVKS  
 IINGCAARDKGLRVNDQILVNGSELGKNQDMETLRSMSTEGKRMQILIVA  
 RISKCNELKSPGSPPELPLETALDRERISISLIGLDESPSRNALSLRM  
 KYOLSPVNMQDDTVLIEDRLVLPPLHSDSSSHDDVGVFTADAGWAARAI  
 SDSACPSLSPVDVPLAFORGEFGROIADERTKLNIVDDOKAGSPRDVPSLGLKSS  
 SLESLQTAVAEYTLNGDIPFRRPRRIIRGRCSNFPRAIDKSTDKRAVDDDDGHE  
 TLEEDTESRSRGRESVSTASDPQSHSLERQNGQEGDKTRKDDTGKRRKDR  
 KEKDMKAKKKGKGLGDMFRGKHKRDKTEKTKIKIOESTEEERIRAKQOER  
 IQAKREFRERARERDVAEIODFRTGCDDELMTYGVSFEYSMALNARQSPREG  
 HMDALVAVOKKPRNSKPSVDSNSRSTPSNHRIOQLROEFOAKODVEDRRYTS  
 FEQPPNARPAATQSGRHSVSVEVOMOROEERSSQOARQYSLPQSRNASSVS  
 ODSWEQNTSPGGRFOSAKENPRYSYQGRNRYLCGHGNARVHLETDLLRQEDRRK  
 EQQMKQPPSEGPSNYSYKVKQDPSVAPKGPFDVPPSPSQVARNLRLOTPEKGR  
 PYS"

misc\_feature 682..942

/gene="PAR3"  
 /note="Region: PDZ domain 1"

misc\_feature 1246..1506

/gene="PAR3"  
 /note="Region: PDZ domain 2"

misc\_feature 1594..1851

/gene="PAR3"  
 /note="Region: PDZ domain 3"

BASE COUNT 1121 a 867 c 1016 g 797 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 500.00 Length: 500  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-757-781-2 x AF252293 ..  
 Align seg 1/1 to: AF252293 from: 1 to: 3801  
 857 ILEAlaSPGLuThrLysLeuAsnThrValAspAspGlnLysAlaGlyse 873  
 2299 ATAGCTACAGACTAACTCAATACAGTGGATGACCGAAAGAGGTTTC 2348  
 873 rProSerArgAspValGlyProSerLeuGlyLeuLysSerSerSert 890  
 2349 TCCAGACAGAGATGGTGGTCCCTTCCGTGGGTCTGAGAGAGTCGTCGT 2398  
 890 euGluSerLeuGlnThrAlaValAlaGluValThrLeuAsnGlyAspIle 906  
 2399 TGGAGAGCTTCGACAGACCGCAGTCCGAGGTGACTTTGAATGGGGATATT 2448  
 907 ProPheHisArgProArgProArgIleIleArgGlyValArgGlyCysAsnGly 923  
 2449 CTTTCCATGCTCCACGCCGCCGGATTAATCAGAGCGAGGGGATCCAATGA 2498  
 923 UserPheArgAlaAlaIleAspLysSerTyrAspLysProAlaValAspA 940  
 2499 GAGCTCAGAGCTCCATCGACAAATCTTATGATAAACCCGCCGTAGATG 2548  
 940 spAspAspGluGlyMetGluThrLeuGlnLysAspThrGluGluSerSer 956  
 2549 ATGATGTAGAGGGCATGAGACCTTGGAGAGACACAGAAAGAAAGTTCA 2598  
 957 ArgSerGlyArgGluSerValSerThrAlaSerAspGlnProSerHisSe 973  
 2599 AGATCAGGAGAGAGGCTGTATCCACAGCCAGTATCAGCTTCCCACTC 2648  
 973 rLeuGluArgGlnMetAsnGlnGlnGluLysGlyAspLysThrAspA 990  
 2649 TCTGGAGAGCAAAATGATGAAACCAAGAGAAAGTGATTAAGACTATA 2698  
 990 rGlyLysAspLysThrGlyLysGluLysLysLysAspArgAspLysGlu 1006  
 2699 GAAAAAAGGATAAATGGAAGAAAAAGAAAGAAATAGAGATPAAGAG 2748  
 1007 LysAspLysMetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMe 1023  
 2749 AAGGATTAATTAAGAAAGCCAGAAAGGAAATGCTGAAGGGCTTGGAGACAT 2798  
 1023 tPheArgPheGlyLysHisArgLysAspAspLysIleGluLysThrGlyL 1040  
 2799 GTTCAGGTTGGCAAAACATCGAAAAAGATGACAAATTCAGAAAAACGGGTA 2848  
 1040 ysiIleLysIleGlnGluSerPheThrSerGlnGluGluValArgMet 1056  
 2849 AATTAATAAAATACAGAAATCCTTACATCAGAAAGAGAGAGATACGAATG 2898  
 1057 LysGlnGlnGlnGluArgIleGlnAlaLysThrArgGluPheArgGluAr 1073  
 2899 AAGCAGGAGGAGAGAGAGATTCAGCCAAACCTGAGAAATTAAGGAACG 2948  
 1073 gGlnAlaArgGluArgAspTyrAlaGluIleGlnAspPheHisArgThrP 1090  
 2949 ACAAGCTCGAAGCGTGAATCTGAAATTCAGAAATTCATTCATCGACAT 2998  
 1090 heGlyCysAspAspGluLeuMetTyrGlyGlyValSerSerTyrGluGly 1106  
 2999 TTGGCTGTGATGATGAGTTAATGTATGGGGAGGATTTCTTCTTATGAAGCT 3048

1107 SerMetAlaLeuAsnAlaArgProGlnSerProArgGluGlyHisMetMe 1123  
 3049 TCCATGGCTCTCAACGCTAGACCTCAGACGCCACAGAGAGGCGATATGAT 3098  
 1123 tAspAlaLeuTyrAlaGlnValLysLysProArgAsnSerLysProSerP 1140  
 3099 GGATGCTTTGTATGCCCAAGTCAGAAAGCCGCCGAATTCCAAAACCTTCAC 3148  
 1140 roValAspSerAsnArgSerThrProSerAsnHisAspArgIleGlnArg 1156  
 3149 CTGTAGACAGTAAACAGATCAACTCTTACAAATCATGATCGGATACAGCT 3198  
 1157 LeuArgGlnGluPheGlnGlnAlaLysGlnAspGlnLysAspValGluAsp 1173  
 3199 CTGAGGCAAGAAATTCAGCAAGCAAGCAAGATGAAGATGTAAGAAATCG 3248  
 1173 gArgArgThrTyrSerPheGlnGlnProTyrProAsnAlaArgProAlaT 1190  
 3249 TCGCGCAGACCTATGATTTTGTAGCAACCTGGCCGACAGCGACGCCGCGA 3298  
 1190 hrgInSerGlyArgHisSerValSerValGluValGlnMetGlnArgGln 1206  
 3299 CGCAGACCGGGCGACACTCGGTGTCGTGGAGGTGCAGATGCACCGGCGAG 3348  
 1207 ArgGlnGlnGluArgGluSerSerGlnGlnAlaGlnArgGlnTyrSerSe 1223  
 3349 CGGAGAGAGAGAGCCCGAGAGCTCCACAGAGCCCAAGCCCGCATACAGCTC 3398  
 1223 rLeuProArgGlnSerArgLysAsnAlaSerSerValSerGlnAspSerT 1240  
 3399 TCTCCTTCGCGCAAGACAGAAATAATGCCAGCTCCGCTCCACAGACTCTT 3448  
 1240 rPglGlnAsnTyrSerProGlyGluGlyPheGlnSerAlaLysGlnAsn 1256  
 3449 GGGAGCAAGACTACTCCCTGGGAAGGCTTCCAGAGTCCCAAGAGAGAAC 3498  
 1257 ProArgTyrSerSerTyrGlnGlySerArgAsnGlyTyrLeuGlyGlyH 1273  
 3499 CCCAGGTACTCTCAGACTACCAAGGCTCCAGGAAGGCTACCTGGGAGGACA 3548  
 1273 sGlyPheAsnAlaArgValMetLeuGluThrGlnGluLeuLeuArgGln 1290  
 3549 TGGCTTCAACCCCAAGGCTCATGTGAAACTCAGAGAGCTCTTCCGAGCG 3598  
 1290 lGlnArgArgLysGlnGlnMetLysLysGlnProProSerGlnGly 1306  
 3599 AACAGAGCGGAGAGAGCGAGATGAAGAAAGACCTCTTCCGAGGGG 3648  
 1307 ProSerAsnTyrAspSerTyrLysLysValGlnAspProSerTyrAlaPr 1323  
 3649 CCCAGCAACTATGACTGTATTAAGAAAGTCCAGAGACCCCAATTACGCCCC 3698  
 1323 oProLysGlyProPheArgGlnAspValProProSerProSerGlnValA 1340  
 3699 TCCCAAGGGCGCTTCCGCGCAAGATGTGCCCTCCCTCTCAGAGTTGG 3748  
 1340 lArgLeuAsnArgLeuGlnThrProGluLysGlyArgProPheArgSer 1356  
 3749 CGAGCTGACAGACTTCAGACTCTGGAAGAGGAGGCCCTTCTATATCC 3798  
 seq\_name: gb\_BC011711  
 seq\_documentation\_block:  
 LOCUS BC011711 1592 bp mRNA linear PRI 02-AUG-2001  
 DEFINITION Homo sapiens, clone MGC:19518 IMAGE:393937.0, mRNA, complete cds.  
 ACCESSION BC011711  
 VERSION BC011711.1 GI:15079815  
 KEYWORDS MGC.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 1592)  
Strausberg, R.  
Direct Submission  
Submitted (30-JUL-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/DTP  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nigrl.nih.gov](mailto:nisc_mgc@nigrl.nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantirpop, S., Thomas, P.,  
Tiongson, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 27 Row: J Column: 9  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7021050.

## FEATURES

## source

1. 1592  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:19518 IMAGE:3939370"  
/tissue.type="Lung, small cell carcinoma"  
/clone\_lib="MIM MGC-7"  
/lab\_host="DH10B-R"  
/note="Vector: POTB7"  
460. 1503  
/codon\_start=1  
/product="unknown (protein for MGC:19518)"  
/protein\_id="AAH1711.1"  
/db\_xref="gi:15079816"  
/translation="MKAKKMKKGLGDMFRGKHKRDKTEKTKIKIQESTSEER  
IRKQDEKRIQAKTRERKQAREKDYAEIOPDRHFRFGDDDELMTGVSYSGLMNLN  
ARPOSEKREHMDALYAQVKKPRNSPSPVDSNRKSTPSNDRORLROFOCAKDED  
VEDRRRTYSEFQDPNAPRATQGRHSVSEVQOMORQDERSSQAOQRYSLEPRO  
SRKNASVSQDSWQNSYSEGFQSAKENRYSYQSGSRNGYLGSGFNARWALETOE  
ILROEORREOKKQPPSEKSPNSYKQVDPYAPKPGPRQDVPSPSOVARLN  
RLQTPKRGPRFS"

## CDS

BASE COUNT 517 a 355 c 439 g 281 t  
ORIGIN

alignment\_scores:                    length:   497  
                  Quality:   497.00                    Gaps:   0  
                  Ratio:   1.000                    Percent identity: 100.000  
Percent similarity: 100.000

alignment\_block:  
US-09-757-781-2 x BC011711 ..

Align seg 1/1 to: BC011711 from: 1 to: 1592

860 GluThrLysLeuAsnThrValAspAspGlnLysAlaGlySerProSerAr 876  
|||||  
10 GAGACTAACTCAATACAGTGATGACCAAGAAAGCAGGTTCTCCACAGCAG 59  
876 gAspValGlyProSerLeuGlyLeuLysLysSerSerSerLeuGluSerL 893  
|||||  
60 AGATGTGGTCTCTCCCTGGGCTCTGAAGAAAGTCAACCTCATTTGAGAGTC 109

893 euglnThrAlaValAlaGluValThrLeuAsnGlyAspIleProPheHis 909  
110 TGCAGACCGCAGAGTTGCCAGAGTACTTGAATGGGATATCTTCTCCAT 159  
910 ArProArProArGilelleArGlyArGlyCysAsnGlySerPheAr 926  
160 CGTCCACGCGCGCGATATCATCAGAGCGAGGATGCAATGAGAGCTTCAG 209  
926 gAlaAlaIleAspLysSerTyrrAspLysProAlaValAspAspAspG 943  
210 AGCTGCATCGCAAAATCTATGATTAACCCCGGAGATGATGATGATG 259  
943 LuGlyMetGluThrLeuGluGluLysPheGluGluGluSerSerArgSerGly 959  
260 AAGGATGAGACCTTGAGACAGACACAGAGAAATTCATCAATGATCAGGG 309  
960 ArgGluSerValSerThrAlaSerAspGlnProSerHisSerLeuGluAr 976  
310 AGAGAGTCTGTATCCACAGCCAGATGATCAGCTCCACCTCTGAGAG 359  
976 gGlnMetAsnGlyAsnGlnGlnGlyGlyAspLysThrAspArgLysLysA 993  
360 ACAATGATGGAACCAAGAGAAAGTGATACACTGATGAGAAAAAGG 409  
993 sPlySerThrGlyLysGluLysLysLysAspArgAspLysGluLysAspLys 1009  
410 ATTAAGCTGGAAGAAAGAAAGAAAGATAGATAGATAGAGAGATTA 459  
1010 MetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArgP 1026  
460 ATGAAAGCCCAAGAGAGAGTGTGAAGGCTTGAGAGACATCTTCAGTT 509  
1026 eGlyLysHisArgLysAspAspLysIleGluLysThrGlyLysIleLysI 1043  
510 TGGCAACATCGAAAGAAAGATGACAGATTCAGAAACGGGTAAATTA 559  
1043 leGlnLysSerPheThrSerGluGluGluArgIleArgMetLysGlnGlu 1059  
560 TACGGAATCTTTACATCAGAGAGAGAGATGAGATGAGTGAAGAGAG 609  
1060 GlnGluArgIleGlnAlaLysThrArgGluPheArgGluArgGlnAlaAr 1076  
610 CAGGAGAGATTCACCAACCAACTCGAATTTAGAGAGCAAGACCTCG 659  
1076 gGluArgAspTyrrAlaGluIleGlnAspPheHisArgThrPheGlyCysA 1093  
660 AGAGCGTGAATGCTGAATTCAGATTTTCATCGACATTTGGCTGTG 709  
1093 sPAspGluLeuMetTyrrGlyGlyValSerSerTyrrGluGlySerMetAla 1109  
710 ATGATGAGTTAATGTATGGGAGGTTCTTCTATGAAGTTCCATGGCT 759  
1110 LeuAsnAlaArgProGlnSerProArgGluGlyHisMetLeuAspAlaLe 1126  
760 CTCAGACCTAGACCTCAGACCCACAGAGGCGCATATGATGATGATCTT 809  
1126 UTyrrAlaGlnValLysLysProArgAsnSerLysProSerProValAspS 1143  
810 GTATGCCCAAGTCAAGAGCGCGGAGATTCCAAACTCATCCTGTAGACA 859  
1143 eTAsnArgSerThrProSerAsnHisAspArgIleGlnArgLeuArgGln 1159  
860 GTACAGATCAACCTCTAGCAATCATGATGAGATACAGCTCTGAGGCAA 909  
1160 GluPheGlnGlnAlaLysGlnAspGluAspValGluAspArgArgArgTh 1176  
910 GAATTTCAAGCAAGCAAGCAAGATGAAGATGATGAAGATGCTCGCGGAC 959  
1176 rTyrrSerPheGluGlnProTrrProAsnAlaArgProAlaThrGlnSerG 1193  
960 CTATAGTTTGAACACCTGCGGAGACGAGCGGCGGCGAGAGAGCG 1009

```

1193 1yatgHisSerValSerValGluValGlnMetGlnArgGlnArgGlnGlu 1209
|||||
1010 GGCACACTCGTGTCCGTGAGGTGCAGATGCAGCGCACCGCAGAGAG 1059
|||||
1210 GluArgGluSerSerGlnGlnAlaGlnArgGlnArgSerSerLeuProAr 1226
|||||
1060 GAGCGCGAGACCTCCACAGAGGCCAGCCAGACAGTACAGCTCTGCTCG 1109
|||||
1226 GglnSerArgLysAnaLaserSerValSerGlnAspSerTrpLuglna 1243
|||||
1110 GCAAAGACAGAAAAATGCAGCTCGGTCTCCAGAGACTCTGGAGACAGA 1159
|||||
1243 snTySerProGlyGluGlyPheGlnSerAlaLysGluAsnProArGlyr 1259
|||||
1160 ACTCTCTCCCTGGGGAAGGCTTCCAGAGTGCAGAAAGAACCCAGGTAC 1209
|||||
1260 SerSerTyArgGlnGlySerArgAsnGlyTyTrpLeuGlyGlnGlyPheAs 1276
|||||
1210 TCCAGCTACCAAGGCTCCAGAGACGCTACCTGGGAGAGACATGGCTCAA 1259
|||||
1276 nAlaArgValMetLeuGluThrGlnGlnLeuLeuArgGlnGlnArgA 1293
|||||
1260 CGCAGAGGTATGCTGGAACACTCAGAGACTCTTCGCCAGAGACAGAGGC 1309
|||||
1293 rGlyGlnGlnGlnMetLysLysGlnProProSerGluGlyProSerAsn 1309
|||||
1310 GGAAGGAGCACAGATGATGAAGACAGCTCTCTCCAGAGGCGCCAGCAGC 1359
|||||
1310 TyraSPSerTyLysValGlnAspProSerTyraLapProProLysAGL 1326
|||||
1360 TATGACTCTGTATTAAGAAAGTCCAGAGACCCAGTTACGCCCTCCCAAGG 1409
|||||
1326 yProPheArgGlnAspValProProSerProSerGlnValAlaArgLeuA 1343
|||||
1410 GCCCTCCGGAAGATGAGCCCCCTCCCTTCAGAGGTGCGAGGCTGA 1459
|||||
1343 snArgLeuGlnThrProGlnLysGlyArgProPheTySer 1356
|||||
1460 ACAGACTTCAGACTCTGAGAAAGGAGGCCCTCTATATTC 1500
|||||
seq_name: gb_pr:AF177228
seq_documentation_block:
LOCUS AF177228 2004 bp mRNA linear PRI 17-JAN-2001
DEFINITION Homo sapiens CTCL tumor antigen seq-5 mRNA, partial cds.
ACCESSION AF177228
VERSION AF177228.1 GI:11275611
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Eichmuller,S., Usener,D., Dummer,R., Stein,A., Thiel,D. and
Schadenorff,D.
TITLE Serological detection of cutaneous T-cell lymphoma-associated
antigens
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (2), 629-634 (2001)
MEDLINE 21143360
REFERENCE
AUTHORS Eichmuller,S., Usener,D., Dellus,H. and Schadenorff,D.
TITLE Direct Submussion
JOURNAL Submitted (10-AUG-1999) Dermato-Oncology (D0900), German Cancer
Research Center, Im Neuenheimer Feld 280, Heidelberg 69120, Germany
FEATURES
source
1..2004
/organism="Homo sapiens"
/db_xref="taxon:9606"
/issue_type="testis"
/note="Identified by a SEREX approach screening a testis
cDNA library using sera from CTCL patients"
<1..>2004
/codon_start=3

```

```

/product="CTCL tumor antigen seq-5"
/protein_id="AA633676.1"
/db_xref="GI:11275612"
/translation="EHEHLEFRFENDCIVIRINDGDLNRNRFEOAHMFARPTIIFH
VYPAANKEDYEDLOSCEKNYYSRSRPSQVYIDNRSVAGLHTYVORAPRIHNPEO
IDSHSLRPHSAHPSCRPSPASARQNVSTYSSYNTKIGKRLNIDLKGTBELE
FSTISRDVITIGSAPITVKNILPRGAIDIDGRLKADRLIEVGDVLGKSOEYVSL
LRSTKMEGVSLVFEQEDAFHPRLEINAPDSQMIPEKRADEDEDIVLPDGTREFLT
FEVPLNDSSGAGLGVSKNGRSKENHADLIGFVKSILINGAASKDGRLEVNQLAVN
GESLIGKTNDAMETLRSMSTENRGMIOILVARIKCNELKSPGPELPIE
TALDDEERLISHSTVIGTIGLDESPRNALTRINGESGKYOLSPVNNPODVTYIE
DDRRLPVLPRLHSDQSSSHDDVGEPTADAGTMAKRAISDSLSIPVDVYLAFOR
EGHROIADETKLVNTYDDQKAGSPSRDVPISGLKSSLESLOQTAIVAVTYLNGDIPF
HRPRRIIRGRCNESPRAIDKSYDKPAVDDDDDEGMETLEEDTDESSRSRGRESVTA
SDPSHSLEROMNGNQEKGDKTDKRRKKKKKKKKKKKKKKKKKKKKKKKKKKKK"
BASE COUNT 647 a 429 c 495 g 433 t
ORIGIN
alignment_scores:
Quality: 427.00 Length: 515
Ratio: 0.831 Gaps: 1
Percent Similarity: 99.806 Percent Identity: 99.806
alignment_block:
us-09-757-781-2 x AF177228
Align seg 1/1 to: AF177228 from: 1 to: 2004
313 GlnHISGluAnLeuPheArgGluAsnAspCysIleValArgIleAsnAs 329
|||||
3 GAAcATGAAAAATCTTTTCGTGAGAAATGATGCAATGTCCAGAGATTAAGA 52
329 pGlyAspLeuArgAsnArgArpPheGlnGlnAlaGlnHisMetPheArg 346
|||||
53 TGCCACCTTTCGAAATAGAAcATTGAAcAAAGcAAcATATGTTGCCc 102
|||||
346 InAlaMetArgThrProIleIleTrpPheHisValAlaProAlaAlaAsn 362
|||||
103 AAGCATGCGTAAcACCCATcATTGTTCCATGTTGTTCTCGACAAAT 152
|||||
363 LysGlnGlnTyArgGlnLeuSerGlnSerGlnLysAsnAsnTyTyTySe 379
|||||
153 AAAGAGcAGTATGAACAcTATCCAAAGTGAAGAAcAAATTAcTATTC 202
|||||
379 rSerArgPheSerProAspSerGlnTyTrIleAspAsnArgSerValAsn 396
|||||
203 AAGCGTTTATGcCTCGACAGCCAcTATATGACAAcAGAGATGTAAC 252
|||||
396 eTAlaGlyLeuHisThrValGlnArgAlaProArgLeuAsnHisProPro 412
|||||
253 GTGCAGGcGCTTCAcAGCGTGAAGAGACACCCGAcTAAcACCGCGCT 302
|||||
413 GlnGlnIleAspSerHisSerArgLeuProHisSerAlaHisProSerG 429
|||||
303 GAGcAGATAGAcTCTcCAcTCAAGAcTAcTAcTAcTAcTAcTAcTAcT 352
|||||
429 LysProProSerAlaProAlaSerAlaProGlnAsnValPheSerThr 446
|||||
353 AAACcACAcTCCGCTCCAGcCTCGGcAcTCAcAGATATATTAcTAcGA 402
|||||
446 hrValSerSerGlyTyArgAsnThrLysLysIleGlyLysArgLeuAsnIle 462
|||||
403 CTGTAGAGcAGTGTATTAACcAAAAAAATATGCAAGGcGCTTAATATc 452
|||||
463 GlnLeuLysLysGlyThrGlnGlyLeuGlyPheSerIleHisSerArgA 479
|||||
453 CAGCTTAAGAAAGGTACAGAGGTTT.GCAATTCAGAcTCAcCTTCcAAG 501
|||||
479 sPValThrIleGlyLysSerAlaProIleTyValLysAsnIleLeuPro 495
|||||
502 ATGTAAcATAGTGTGCTCAGCTCCAAATCTATGTGAAAAAcATTCTCCc 551
|||||

```

```

496  ARGGLAAlaAlaIleGlnAspGlyArgLeuYsAlaGlyAspArgLeuI 512
    |||||||
552  CGGGGGGGCCCATTCAGATGGCCGACTTAAGCAGAGACAGACTTAT 601
    |||||||
512  eglValaanglyValaAspLeuValGlyYsSerGlnGluValaValS 529
    |||||||
602  AGAGGTAAATGGAGTATTAAGTGGGCAATCCCAAGAGAGAGTGT 651
    |||||||
529  erleuLeuArgSerThrIysMetGluGlyThrValSerLeuLeuValPhe 545
    |||||||
652  CGCTGTTGAGAACACCAAGATGAGAGAACTGTAGCTTGTGTCCTT 701
    |||||||
546  ArgGlnGluAspAlaPheHisProArgGluLeuAsnAlaGluProSerG 562
    |||||||
702  CGCCAGAGAGACGGCTTCCACCAAGGAGATTAATGACAGACCAAGCA 751
    |||||||
562  nMeGlnIleProIysGluThrIysAlaGluAspGluAspIleValLeu 579
    |||||||
752  GATCAGATTCACAAAGAAAGCAAGACAGAAAGTGAAGATATGTTCT 801
    |||||||
579  hrProAspGlyThrArgGluPheLeuThrPheGluValProLeuAsnAsp 595
    |||||||
802  CACCTGATGGCAGCAGGAAATTTCTGACATTGAAAGTCCACACTTAAT 851
    |||||||
596  SerGlySerAlaGlyLeuGlyValaSerValIysGlyAsnArgSerIysG 612
    |||||||
852  TCAGATATTCAGAGGCTTGTGTGTCAGTGTCAAGGTAAACGGTCAAA 901
    |||||||
612  uAsnHisAlaAspLeuGlyIlePheValIysSerIleIleAsnGlyGly 629
    |||||||
902  GAACACAGCAGATTTGGGAATCTTGTCAACATTCATTAATATGAGAG 951
    |||||||
629  IaAlaSerIysAspGlyArgLeuArgValaAsnAspGlnLeuIleAla 645
    |||||||
952  CAGACATCAAAAGATGAGAGGCTTCGGGTGAATGATCACTGATAGCA 1001
    |||||||
646  AsnGlyIysLeuLeuGlyIysThrAsnGlnAspAlaMetGluThrLe 662
    |||||||
1002  AATGGAGAAATCCCTGTGGGCAAGCAAAACCAAGATGCCATGGAAAC 1051
    |||||||
662  uArgAspSerMetSerThrGlyIleAsnIysArgGlyMetIleGlnLeu 679
    |||||||
1052  AAGAGGTCTAATCTTACTGAAAGGCAATTAACAGAGAAATGATCCAG 1101
    |||||||
679  IeValAlaArgArgIleSerIysCysAsnGluLeuLeuSerProGlySer 695
    |||||||
1102  TTGTTCGAAGAGAAATAGCAAGTGAATGAGTGAAGTCACTGGAGAC 1151
    |||||||
696  ProProGlyProGluLeuProIleGluThrAlaLeuAspAspArgGlu 712
    |||||||
1152  CCCCCTGACCTGAGCTGCCCATTTGAACAGCGCTGATGATGAGAGAC 1201
    |||||||
712  gArgIleSerHisSerLeuIysSerGlyIleGluGlyLeuAspGluSerP 729
    |||||||
1202  AAGAAATTTCCCATTCCTCTACAGTGGGATGAGGGCTTGATGAATCC 1251
    |||||||
729  roSerArgAsnAlaAlaLeuSerArgIleMetGlyIysSerGlyIysTyr 745
    |||||||
1252  CCAGAGAAATGCTGCCCTCAGTGAATGGGTGAGTCAAGTAAATAC 1301
    |||||||
746  GlnLeuSerProThrValaAsnMetProGlnAspAspThrValIleIle 762
    |||||||
1302  CAGCTGTCCTACAGTGAATATGCCCAAGATACACACTGATTAATAGA 1351
    |||||||
762  uAspAspArgLeuProValaLeuProProHisLeuSerAspGlnSerS 779
    |||||||
1352  AAGATACAGGTGGCCAGTGTCTCCCAATCTCTGACACAGTCCACTT 1401
    |||||||
779  erSerSerHisAspAspValaGlyPheValThrAlaAspAlaGlyThrT 795
    |||||||
1402  CCAGGCTCCATGATGATGTGGGGTTGTGACGGCAGATGCTGGACTT 1451
    |||||||
796  AlaIysAlaAlaIleSerAspSerAlaAspCysSerLeuSerProAsp 812
    |||||||

```

seq name: gp\_pat:BD004905

seq documentation block:

LOCUS BD004905 2718 bp DNA linear PAT 31-JAN-2002

DEFINITION Novel genes encoding protein kinase or protein phosphatase.

ACCESSION BD004905

VERSION BD004905.1 GI:18632866

KEYWORDS JP 03074935-T/2.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,

Funahashi,S., Sano,C. and Nezu,J.

Novel genes encoding protein kinase or protein phosphatase

Patent: JP 03074935-T 2 30-JAN-2001;

Novel genes encoding protein kinase or protein phosphatase

HELIIX RESEARCH INSTITUTE, TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA,

KOJI HAYASHI, KAOH SAITO, JUNICHI YAMAMOTO, SHIZUKO ISHII, OMOKYASU

SUGIYAMA, AI WAKAMATSU, KEIICHI NAGAI, TETSUJI OTSUKI, SHINICHI

FUNAHASHI, HIKKI SENO, JUNICHI NEZU

OS Homo sapiens (human)

PN JP 03074935-T/2

PD 30-JAN-2001

PF 28-JUL-2000 JP 2000005061

PR 29-JUL-1999 JP 99P 248036, 11-JAN-2000 JP 00P 118776 PR

02-MAY-2000 JP 00P 183767, 09-JUN-2000 JP 00P 241899 PR

18-OCT-1999 US 60/159590, 17-FEB-2000 US 60/183322 PI TOSHIO

OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, PI KAOH SAITO,

PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,

PI KEIICHI NAGAI, TETSUJI OTSUKI, SHINICHI FUNAHASHI, CHIKI SENO,

PI JUNICHI NEZU

PC C12N15/12, C12N9/12, C12N9/16, C12N1/15, C12N1/19, C12N1/21, PC

C12N5/10, C07K16/40,

PC C1201/68, C1201/42, C1201/48

CC

FH

FT

FEATURES

source

BASE COUNT 840 a 586 c 688 g 604 t

ORIGIN

alignment scores:

Quality: 420.00 Length: 508

Ratio: 0.830 Gaps: 1

Percent Similarity: 99.606 Percent Identity: 99.606

alignment block:

US-09-757-781-2 x BD004905

Align seg 1/1 to: BD004905 from: 1 to: 2718

234 GtlyGlnGlnGlnGlnAspGlyIleThrGluGlnAspAsnSerArgVal 250

|||||

327 GAGAAACAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTCT 376

|||||

250 IgluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPheS 267

|||||

377 TGACCTGTGTGACACTGCTGACACGGGTTTGACACATATACCCAACTTT 426

|||||

267 erleuAspAspMetValIysLeuValGluValProAsnAspGlyGlyPro 283

```

|||||
427 CTCTGGATGATATGTAAGCTCTAGAAAGTCCCAACGATGAGGGCT 476
284 LeuGlyIleHsValValProPheSerAlaArgGlyArgThrLeuG1 300
|||||
477 CTGGATTCCTATAGTCCCTTCAGTCCGAGAGCGGAGAACCTGG 526
300 YLeuLeuValLysArgLeuGlyGlyGlyLysAlaGluHsGluSnl 317
|||||
527 GTATTAGTAAACGATTGAGAAAGGTGTAAGCTGAACATGAAATC 576
317 eupheaArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 333
|||||
577 TTTTTCGAGAAATGATTCATTCAGATTAATGATGCGACCTTCGA 626
334 AsnArgArgPheGluGluAlaGluHsPheArgGluAlaMetArgTh 350
|||||
627 AATGAAAGATTGAAACAAACCAATATGTTTCCCAAGCCATGCGTAC 676
350 rProIleIleThrPheHsValValProAlaAlaAsnLysGluInTyrg 367
|||||
677 ACCCATATTGTTGTTCCATGTGTTCCGACAAATTAAGACAGTATG 726
367 LueGluLeuSerGluLysAsnAsnTyrrTyrrSerSerArgPheSer 383
|||||
727 AACAACTATCCCAAGTGAAGAAACAATTAATTCATTCAGCCGTTTAA 776
384 ProAspSerGlnTyrrIleAspAsnArgSerValAsnSerAlaGlyLeuH1 400
|||||
777 CCGACACCCAGCTATATGACACAGAGGTGTAACAGTGCAGGGCTTCA 826
400 sThValGlnArgAlaProArgLeuAsnHsProProGluGlnIleAspS 417
|||||
827 CACGCTGACAGAGACACCCCACTGAACACCCGCTGACGACAGTATGACT 876
417 eHsIleSerArgLeuProHsSerIleHsProSerGlyLysProProSer 433
|||||
877 CTCACCTACAGACTCTCTATAGCGACACCCCTGGGAAACCCACATCC 926
434 AlaProAlaSerAlaProGluAsnValPheSerThrThrValSerSerG1 450
|||||
927 GCTCAGGCTCGGACCTCAGATGTATTTAGTACGATGTAAAGACAGTGG 976
450 YTrAsnThrLysLysIleGlyLysArgLeuAsnIleGluLeuLysLysG 467
|||||
977 TTATATACACCAAAAAATAGGCAAGAGCTTAATATCCAGCTTAAGAAAG 1026
467 LyrThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 483
|||||
1027 GTACGAAAGGTTTGGGATTCAGCATCTCCAGAGATGTAAACAATVAGCT 1076
484 GlySerAlaProIleTyrrValLysAsnIleLeuProArgGlyAlaAla11 500
|||||
1077 GGCTAGCTCCCATCTATGTGAAAAACATCTCCCGGGGGGGCCCAT 1126
500 eGluAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyV 517
|||||
1127 TCAGATATGCCGACTTAAGCAGACAGACAGACTTAAGAGTTAAATGAGAG 1176
517 AlaAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSer 533
|||||
1177 TAGATTGTGTGGCAAAATCCCAAGAGAAAGTTGTTCCCTGTGGAAGGC 1226
534 ThrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAl 550
|||||
1227 ACCAAGATGGAAGAACTGTGAGCTTCTGCTTCCGCCAGAGAGACC 1276
550 aPheHsProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProL 567
|||||
1277 CTTCACCCCAAGGAACTGAATGACAGACCAAGCCAGATGAGATTCCAA 1326
567 yGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThr 583
|||||

```

```

1327 AAGAAAGCAAAAGCAGAAAGATGAGATATGTTCTTTACACCTGATGGCACC 1376
584 ArgGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaG 600
|||||
1377 AGGGAATTTCTGACATTTGAAAGTCCCACTTAGT.GATTACAGATCTGGAG 1425
600 LysLeuGlyValSerValLysGlyAsnArgSerLysGluAsnHsAlaAsp 616
|||||
1426 GCCTTGGGTGTCAGTCCAAAGTAAACCGTCAAAAGAGAACACGACAGAT 1475
617 LeuGlyIlePheValLysSerIleIleAsnGlyLysAlaAlaSerLysAs 633
|||||
1476 TTGGGAATCTTTGTCAAGTCCATTATTAATGACGACGACATCTTAAGA 1525
633 pGlyArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerL 650
|||||
1526 TGGAAGGCTTGGGGATGATGATCAACTGATGACGATAATGAGATATCC 1575
650 euleGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMet 666
|||||
1576 TGTGGGCAAGCAAAACCAAGATGCCATGGAACCTTAAGAAAGTCTATG 1625
667 SerThrGluGlyAsnLysArgGlyMetIleGluLeuIleValAlaArgAr 683
|||||
1626 TCTACTGAAGCCAAATAACGAGAAATGATCCAGCTTATTTGTCAGAAAG 1675
683 gIleSerLysCysAsnGluLeuLysSerProGlySerProProGlyProg 700
|||||
1676 AATAAGCAAGTGCAATGAGTGAAGTACCTGGGAGCCCTTGACACTG 1725
700 LuLeuProIleGluThrAlaLeuAspAspArgGluArgIleSerHis 716
|||||
1726 AGCTGCCATTTGAAACACCGTTGATGTATGAGAAAGCAAGAAATTTCCAT 1775
717 SerLeuTyrrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAl 733
|||||
1776 TCCTCTACAGTGGGATTTGAGGGGCTTGAATGATGCCACGACGAAGTCC 1825
733 AlaLeuSerArgIleMetGly 740
|||||
1826 TGCCCTCAGTAGGATTAATGGGT 1847
seq_name: gb_pr:AK027735
seq_documentation_block:
LOCUS AK027735 2718 bp mRNA linear PRI 15-MAY-2001
DEFINITION Homo sapiens cDNA FLJ14829 fis. clone OVARC100945, moderately
similar to Rattus norvegicus mRNA for atypical PKC specific binding
protein.
ACCESSION AK027735.1 GI:14042633
VERSION AK027735
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens ovary, tumor tissue cDNA to mRNA, clone_11b:OVARC1
clone:OVARC100945.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sities)
REFERENCE
AUTHORS Isegai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Kagatsuna,M., Hosotani,T., Kaku,T., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Iehli,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahari,K., Masuhp,Y., Ninomiya,K. and Iwayanagi,T.
MEDO human cDNA sequencing project
TITLE
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2718)
AUTHORS Isegai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2001) Takao Isegai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-3951, Fax:81-438-52-3952)

```





```

633 pglAargLeuArgValAsnAspGlnLeuIleAlaValAsnGlySerL 650
|||||
1526 TGGAGGCTTCGCGCTGATGATCACTGATAGCAGTAATGAGAGATCC 1575
|||||
650 euleuGlyThrAsnGlnAspAlaMetGluThrIleArgArgSerMet 666
|||||
1576 TGTGGGGAGACAAACCAAGATGCGCATGGAACCCCTAAGAGGCTATG 1625
|||||
667 SerThrGluGlyAsnArgGlyMetIleGlnLeuIleAlaValArg 683
|||||
1526 TCTACTGAGGCAATTAACGAGATGATCCAGCTTATGTTGCAAGAG 1675
|||||
683 gllSerLysCysAsnGluLeuLysSerProGlySerProGlyProG 700
|||||
1676 AATAAGCAAGCTCAATGAGTCAAGTCACTGAGGCCCTCGAGACTG 1725
|||||
700 luleuProIleGluThrAlaLeuAspAspArgGluArgIleSerHis 716
|||||
1726 AGCTGACCATTAAGACAGCGTTGATGATGAGAACGAGAAATTTCCAT 1775
|||||
717 SerLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsn 733
|||||
1776 TCCCTCAAGATGGATGGAGGCTTGATCAATCGCCACAGAAATGC 1825
|||||
733 aAlaLeuSerArgIleMetGly 740
|||||
1826 TGCCCTCAGTAGATTAATGGCT 1847
|||||
seq_name: gb_P1:AK000761

seq_documentation_block:
LOCUS AK000761 2696 bp mRNA linear PRI 22-FEB-2000
DEFINITION Homo sapiens CDNA FLJ20754 fls, clone HEP02246.
ACCESSION AK000761
VERSION AK000761.1 GI:7021050
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HepG2 CDNA to mRNA, clone_1lb:HEP
clone:HEP02246.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (sites)
Kawakami,T., Noguchi,S., Itoh,T., Shigetsu,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human CDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2696)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT
NEDO human CDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; CDNA full insert
sequencing; Research Association for Biotechnology; CDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1. 2696
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP02246"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_1lb="HEP"
/notes="cloning vector pME18SFL3"
45..1088
CDS

```

```

/notes="unnamed protein product"
/codon_start=1
/protein_id="BA91366.1"
/db_xref="GI:7021051"
/translation="MKAKGKMGKGLGDMRPFGRKRRDKIEKTKIKIOESFTSEER
IRKROBERIOATREFREKAREIDYAEIÖDHRIFGDDDELTGVSSTYEGSMALN
ARKPOSREGIMDIALYAQVKKPRNSRPSVDNRSPSNHDIRLÖRLÖFOQKÖDÖD
VEDRRRTYSFEPQPNRAPATÖGRHSVSVEYQMRÖÖERESÖÖAÖROYSSLPÖ
SRKNASSVSÖDÖSEÖNYSPÖEGFÖSÄKENDRYSÖYÖGRNGYLGÖGÖFNARVYLÖEÖ
LÖRÖÖRÖRÖÖKÖKÖPÖSPÖSGÖPÖSNÖYÖSKYÖDPÖSPÖAPKÖPÖRÖDÖVPPÖSPÖVÄRLN
RLÖTPÖKÖRÖPÖFS"
BASE COUNT 703 a 611 c 695 g 687 t
ORIGIN

alignment_scores:
Quality: 356.00 Length: 356
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
us-09-757-781-2 x AK000761
..
Align seg 1/1 to: AK000761 from: 1 to: 2696

1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
|||||
18 AAGATAGAGATTAAGCAGACAGATTAATGAAGCCAGGAAGGAGATGCT 67
|||||
1017 ULysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAspL 1034
|||||
68 GAAGGCTTGGAGACATGTTCAAGCTTGGCCAAATGAAAAAGATGACA 117
|||||
1034 YSILegLysThrGlyLysIleLysIleGlnGluSerPheThrSerGlu 1050
|||||
118 AGATTGACAAAACGGGTAAATTAATTAACAGAAATCTTTRACATCAGAA 167
|||||
1051 GluGluArgIleArgMetLysGlnGlnGlnGluArgIleGlnAlaLysTh 1067
|||||
168 GAGGAGAGATACGAATGACAGAGACGACGAGAGATTCACAGCCAAAC 217
|||||
1067 PArgGluPheArgGluArgGlnAlaArgGluArgAspTyrAlaGluIleG 1084
|||||
218 TCGAATAATTAGGAGACACAAAGCTCGAGAGGCTACATGCTGTAATTC 267
|||||
1084 LAspPheHisArgThrPheGlyCysAspAspGluLeuMetTyrGly 1100
|||||
268 AAGATTTCATCGGACATTTGGCTGTGATGATGATTAATGATGGGGA 317
|||||
1101 ValSerSerTyrGluGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117
|||||
318 GTTCTCTCTTAAGAGGTTCATGCTCAACCGCTACACCTCAAGCC 367
|||||
1117 CArgGluGlyHisMetMetAspAlaLeuTyrAlaGlnValLysLysProA 1134
|||||
368 ACGAATAAGGCAATAGATGATGCTTTGTATGCCAACAAATCAAGAAGCCGC 417
|||||
1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
|||||
418 GGAATTCGAACCCCTCACCTGTAGACAGTACAGATCACTCTTACGAAAT 467
|||||
1151 HisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAs 1167
|||||
468 CATGATCGGATACAGCGCTCGAGCAAGAAATTTCAAGCAAGCAAGACA 517
|||||
1167 pGluAspValGluAspArgArgThrTyrSerPheGluGlnProTyrPP 1184
|||||
518 TGAAGATGTAGAAAGATCGTCGGCGGACCTATAGTTTGAAGCAACCTGCG 567
|||||
1184 rÖAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200
|||||
568 CGAAGCAGCGGCGGCGAGCAGAGACGGGCGACACTGCTGCGTGGAG 617
|||||

```





```

53  TGCACACAGAGAGTGTGAACAGTGCAGGCTTCACACCGTCGACAGAGACAC 102
406  roArgrLeuAnHisProProGluGlnIleAspSerHisSerArgLeuPro 422
103  CCCGACTGAACCCACCGGCTTACAGAGATGAGCTCTCCTACTACAGATACCT 152
423  HtSerAlaHisProSerGlyLysProProSerAlaProAlaSerAlaPr 439
153  CATACCGACACACCCCTCGGGAACACATCCGCTCCAGCCCTCGGACCC 202
439  OGLnAsnValPheSerThrValSerSerGlyTyrAsnThrLysLysI 456
203  TCACAAATGTATTATGTACGACTAGCAGTACGAGTGTATTAACACCAAAAA 252
456  legLyAsrGleuAsnIleGlnLeuLysGlyThrGlnGlyLeuGly 472
253  TAGGCAAGAGGCTTAATATCCACCTTAAGAAAGTACAGAGGTTTGGA 302
473  PheSerIleThrSerArgAspValThrIleGlySerAlaProIleTy 489
303  TTCAGATCACTTCCAGAGATGTATACATAGGTGGCTAGCTCCAACTTA 352
489  rValLysAsnIleLeuProArgrGlyAlaAlaIleGlnAspGlyArgLeu 506
353  TGTGAAACCATTTCTCCCGGGGGGGCCCATTCAGAGTGGCGACTTA 522
506  ySAlaGlyAspArgLeuIleGlnValAsnGlyValAspLeuValGlyLys 522
403  AGGCGAGAGACAGACTTATAGAGTAAATGAGTATGATTTAGTGGCAAA 452
523  SerGlnGluGlnValValSerLeuLeuArgSerThrLysMetGlnGly 539
453  TCCCAAGAGGAAGTGTGCTTGCCTGTGAGAAAGCACCAAGATGAGAAAC 502
539  rValSerLeuLeuValPheArgrGlnGlnAspAlaPheHisProArgrGlu 556
503  TGTGACCTTCTGTGCTTTTCCGAGAGACAGCCCTTCCACCCAGAGGAC 552
556  eudAsnAlaGluProSerGlnMetGlnIleProLysGluThrLysAlaGlu 572
553  TGAATGAGAGCCACGACGATGCAATTCACAAAGAAACGAAAGACAGA 602
573  AspGluAspIleValLeuThrProAspGlyThrArgLysPheLeuThrP 589
603  GATGAGATATGCTTCTTACACCTGATGACACAGGGAATTTCTGACATT 652
589  eGluValProLeuAsnAspSerGlySerAlaGlyLeuGlyValSerVal 606
653  TGAATGCCACTTATGATTCAGATCTGCAAGGCTGTGTGCTAGTCTCA 702
606  ySgLyAsnArgSerLysGlnAsnHisAlaAspLeuGlyIlePheValLys 622
703  AAGGTAACCGGTCAAAAGAGAACCCAGCATCTGGGAATCTTTGTCAAG 752
623  SerIleIle 625
753  TCCATTATT 761

```

seq\_name: gb\_est2:BE792557

seq\_documentation\_block:

LOCUS BE792557 725 bp mRNA linear EST 20-SEP-2000

DEFINITION 60158334.1F1 NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:3939370 5',

ACCESSION BE792557

VERSION BE792557.1 GI:10213755

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 725)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.

```

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at: image.lnl.gov
Plate: ILCH/88 row: b column: 11
High quality sequence stop: 725.
Location/Qualifiers
1..725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3939370"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pOTB7; Site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCGAGAG(5). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 256 a 118 c 207 g 144 t

ORIGIN

alignment\_scores:

Quality: 225.00 Length: 225

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-757-781-2 x BE792557 ..

Align seg 1/1 to: BE792557 from: 1 to: 725

```

867  AspAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuG 883
3  GATGACGAGAAAGAGAGTTCCTCCAGAGAGATGTGGTCTTCCCTCGG 52
883  yLeuLysLysSerSerSerLeuGluSerLeuGlnThrAlaValAlaGlu 900
53  TCGAAGAAAGTCAACCTCATTTGAGAGTCTCCAGACCGGAGTCCGAG 102
900  aLThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIle 916
103  TGACTTTGATGGGATATTCCTTCCATCTCCACGCGCGCGGATATTC 152
917  ArgGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTy 933
153  AGAGGAGAGGGATGCAATGAGACTTCAGAGCTGCCCATGACAAATCTTA 202
933  rAspLysProAlaValAspAspAspAspGluGlyMetGluThrLeuGluG 950
203  TGAATAAACCCCGGTAGATGATGATGATGATGATGATGATGATGATGAT 252
950  LuAspThrGluGluSerSerArgSerGlyArgGluSerValSerThrAla 966
253  AAGACACAGAGAAAGTTCAGATCAGAGAGAGAGTCTGATCCACAGCC 302
967  SerAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlyAsnGlnG 983
303  AGTGATCAGCCTTCCCATCTCTGTGAGAGACAAATGATGAAACCAAG 352
983  uLysGlyAspLysThrAspArgLysLysAspLysThrGlyLysGlnLysL 1000

```

```

353 GAAAGCGATGAAGCTGATGAAAAAGATTAAGTGAAGAAAGA 402
1000 ysltysparasplysglulysasplysmetylalalyslysglymet 1016
403 AGAAGATTAGATTAGAGACAGATTAATGAAGCCAGAAAGGAAATG 452
1017 leuylsGlyleuGlyaspmetpheatrgpheGlylyshisatrglysaspas 1033
453 CTGAAGGGCTTGGGACATGTCAGGTTGGCAAAACATCGAAAAAGATGA 502
1033 plysllleGlylthrglylystlelyslleGlnGlyserPhehrberg 1050
503 CAAGATTGAAGAAACGGGTAAATAAATAACAGAAATCCTTACATCAG 552
1050 luGlnGluAArglleArgmetllysglnGlnGlnGluAArglleGlnAlaLys 1066
553 AAGAGGAGAGATACGATGATGACAGACAGACAGAGAGATTCAGCCCAA 602
1067 ThrArgGluPheArgGluArgGlnAlaAArgGluAArgAspTyrAlaGlu 1083
603 ACTCGAATTTAGGGAACAGCAAGCTCGAGAGCGTGACTATGCTGAAT 652
1083 eGlnAspPheHisArgThrPheGly 1091
653 TCAAGATTTCATCGACATTTGGT 677
seq_name: gb_est2:BM466811

```

```

seq_documentation_block:
LOCUS BM466811 1150 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6431318 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5503072
5', mRNA sequence.
ACCESSION BM466811
VERSION BM466811.1 GI:18515853
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1150)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1AM12142 row: d column: 17
High quality sequence start: 9
High quality sequence stop: 652.
Location/Qualifiers

```

```

FEATURES
source
1. 1150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5503072"
/clone_1lb="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-Sport6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies. 309 g 235 t 2 others

```

```

BASE COUNT 294 a 310 c 309 g 235 t 2 others
ORIGIN

```

```

alignment_scores:
Quality: 213.00 Length: 213

```

```

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
us-09-757-781-2 x BM466811 ..

```

```

Align seg 1/1 to: BM466811 from: 1 to: 1150

```

```

1112 AlArpProGlnSerProArGluGlyHisMetelaspAlaLeuTyrAl 1128
39 GCTGACCTCGAGACCCAGGAAAGGCGATATGATGATGCTTGTATGC 88
1128 aGlnValLysLysProArGlnSerLysProSerProValAspSerAsn 1145
89 TCAGTCAAGAAAGCGCGGAATTCGAAACCTCACTGATACAGTACA 138
1145 rGserThrProSerAsnHisAspArglleGlnAArgleuAArgGlnPhe 1161
139 GATCAACTCCTAGCAATCATGATCGATACAGCCGCTGAGCCAGAAATT 188
1162 GlnGlnAlaLysGlnAspGluAspValGluAspArgArgThrTyrSe 1178
189 CAGCAAGCAAGCAAGATGAAATGATGATGATGATGATGATGATG 238
1178 rPheGlnGlnProTTPProAsnAlaArgProAlaThrGlnSerGlyArgH 1195
239 TTTGAGCAACCTGGCGGAGCGAGCGCGCGCGAGCGAGCGCGGAC 288
1195 lsservalserValGluValGlnMetGlnAArgGlnAArgGlnGluArg 1211
289 ACTGGGTGCTCGTGAGGTGAGATGAGATGAGATGAGATGAGATG 338
1212 GluSerSerGlnGlnAlaGlnAArgGlnTyrSerSerLysProArgGlnSe 1228
339 GAGAGCTCCAGCAGAGCCAGCCAGCCAGCTTCTGCTCGGCAAG 388
1228 rArgLysAsnAlaSerSerValSerGlnAspSerTTPGlnAsnTyrS 1245
389 CAGGAAATGCGCAGCTCGTCTCCAGACTCTGGAGAGCAAGACTCT 438
1245 erProGlyGlyGlyPheGlnSerAlaLysGlnAsnProArgTyrSerSer 1261
439 CCCCTGGGGAAGGCTTCAGAGTGCAGAAAGAACCCAGGACTCCAGC 488
1262 TyrGlnGlySerArgAsnGlyTyrLeuGlyHisGlyPheAsnAlaAr 1278
489 TACCAAGCTCCAGAGACGCTACCTGGAGAGACTTGCTTCACAGCCAG 538
1278 gValMetLeuGlnThrGlnGlnLeuLeuAArgGlnGlnAArgArgLysG 1295
539 GGTGATGCTGGAACCTCAGAGCTCTTCCGAGAGAACAGAGCGGAG 588
1295 luGlnGlnMetLysLysGlnProProSerGlnGlyProSerAsnTyrAsp 1311
589 AGCAGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
1312 SerTyrLysLysValGlnAspProSerTyrAlaProPro 1324
639 TCGTATTAAGAAAGTCCAGAGACCCAGTTACGCCCTCTCC 677
seq_name: gb_est1:AM780417
seq_documentation_block:
LOCUS AM780417 599 bp mRNA linear EST 12-MAY-2000
DEFINITION ho21808.x1 NCI CGAP_C014 Homo sapiens cDNA clone IMAGE:308007 3'
similar to TR:092340 Q92340 AYPICAL PKC SPECIFIC BINDING PROTEIN.
// mRNA sequence.
ACCESSION AM780417
VERSION AM780417.1 GI:7795020
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Thu Jul 25 08:38:28 2002

us-09-757-781-2.0lig.rst

Page 4

REFERENCE 1 (bases 1 to 599)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap/  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapb@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmett-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL, send email to:  
info@image.llnl.gov  
Possible reversed clone: polyA not found  
Seq primer: -400P from Gibco  
High quality sequence stop: 412.  
Location/Qualifiers  
1..599  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:3038007"  
/clone\_lib="NCI-CGAP Col4"  
/tissue\_type="moderately-differentiated adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: SalI;  
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.7 Kb. Life technologies catalog #:  
11531-019"  
BASE COUNT 130 a 157 c 117 g 195 t  
ORIGIN  
alignment\_scores: Quality: 189.00 Length: 189  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent identity: 100.000  
alignment\_block:  
US-09-757-781-2 x AW780417/rev ..  
Align seg 1/1 to reverse of: AW780417 from: 1 to: 599  
808 leuserProaspValaspProvalileuAlaphcGlnArgGluGlypHeG 824  
567 TTGAGTCAGATGTGATCCAGTCTTCCACGAGAGAGATTGG 518  
824 YARGInserMetSerGluysArgThrlyGlnPheserAspAlaSerG 841  
517 ACGTCAGAGTATGTCAGAAAACCCACAAACATTTTCAGATGCCAGT 468  
841 InleuAspPheValysThrArglySerlySerMetAspLeuGlyIle 857  
467 AATTGATTTCGTTAAACAGAAAATCCAAAACCATGGATTAGGATA 418  
858 AlaaspGluThrlyLeuAlaphcGlnArgGluGlypHeG 874  
417 GGTGACGAGACTAACTCAATACAGTGCATGACCAAGAAAGAGTTCC 368  
874 oSerArgaspValGlyProSerleuGlyLeuAlaphcGlnArgGluGly 891  
367 CACGACAGATGTGCTTCCTCGGCTCGAAGAGTCAGAGCTCATTTGG 318  
891 InSerleuGlnThrAlaValAlaGluValThrleuAsnGlyAspIlePro 907  
317 AGAGTCTGCAGACCGCAGTGGCCGAGGACTTGAATGGGATATTCCT 268  
908 PheHisArgProArgProArgIleIleArgGlyArgGlyCysAsnGlu 924  
267 TTCATGTCTCAGCGCCGAGTAAATCAGAGCGAGGATGCAATGANG 218

924 rPheArgAlaAlaIleAspLysSerlyrAspLysProAlaValaAspA 941  
117 CTTCAGAGCTCCCATCGACAAATCTTATGATTAACCGCGGTGATGAG 168  
217 CTTCAGAGCTCCCATCGACAAATCTTATGATTAACCGCGGTGATGAG 168  
941 SPASpGluGlyMetGluThrleuGluGluAspThrGluGluSerArg 957  
167 ATGATGAAGCAGTGGACACCTTGAGAAAGACACAGAAAGAAATTCAGA 118  
958 SerGlyArgGluSerValSerThrAlaSerAspGlnProSerHisSer 974  
117 TCAGGAGAGAGACTGTATCCACAGCCGATATCAGCTTCACACTTCCT 68  
974 uGluArgGlnMetAsnGlyAsnGluGluGlyAspLysThrAspArg 991  
67 GCGAGACAAATGATGAGAAACCAAGAGAAAGTGATTAAGCTATGAA 18  
991 YSLysAspLysThrGly 996  
17 AAAAGATTAACACTGA 1  
seq\_name: gb\_est2:BE693704  
seq\_documentation\_block:  
LOCUS BE693704 645 bp mRNA linear EST 11-SEP-2000  
DEFINITION RC0-BT0387-100800-013-f05 BT0387 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE693704  
VERSION BE693704.1 GI:10080864  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Mammalia: Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 645)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.C.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=62=RC0-BT0387-100  
800-013-f05&l3=2000-08-10&l4=1)  
800-013-f05&l3=2000-08-10&l4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 45  
High quality sequence stop: 644.  
Location/Qualifiers  
1..645  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BT0387"  
/dev\_stage="Adult"  
/note="Organ: breast; Vector: puc18; Site: 1: SmaI; Site: 2:  
SmaI; A mini-library was made by cloning products derived  
from ORFESTS PCR (U.S. Letters Patent application No. 196  
716 - Ludwig Institute for Cancer Research) profiles  
into the puc 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."  
BASE COUNT 122 a 176 c 115 g 232 t  
ORIGIN

alignment\_scores:                      Quality: 188.00                      Length: 188  
   Ratio: 1.000                      Gaps: 0  
   Percent Similarity: 100.000           Percent Identity: 100.000

alignment\_block:  
 US-09-757-781-2 x BE693704/rev ..

Align seg 1/1 to reverse of: BE693704 from: 1 to: 645

```

838 ASPALASerGlnLeuAspPheValLysThrArgLysSerIysSerMetas 854
|||||
610 GATGCCAGTCATATGGATTTCGTTAAACAGAAAATCAAAAGACATGGA 561
854 PleaGlyIleAlaAspGluThrLysLeuAsnThrValAspAspGlnLysA 871
|||||
560 TTTAGGATAGCTGACGACGACTAACTCAATACAGTGATGACCAAGAAAG 511
871 laGlySerProSerArgAspValGlyProSerLeuGlyLeuLysLysSer 887
|||||
510 CAGGTTCTCCACAGCAGAGATGGGTCTCCCTGGGTCTGAGAAACTCA 461
888 SerSerLeuGlnSerLeuGlnThrAlaValAlaGluValThrLeuAsnG 904
|||||
460 AGCTCATTTGAGAGTCTGCAGACCGCAGTTGCCGAGGTGACTTGAATG 411
904 YAspIleProPheHisArgProArgProArgIleIleArgGlyArgGly 921
|||||
410 GGATATTCCTTCATCCTCCACGCGCGCGGATATACAGAGCGGCGGAT 361
921 YAsnGlnSerPheArgAlaAlaIleAspLysSerIysAspLysProAla 937
|||||
360 GCATGAGAGCTTCAGAGCTGCCATCGACAATCTTATGATTAACCCGCG 311
938 ValAspAspAspAspGluGlyMetGluThrLeuGlnGluAspThrGlu 954
|||||
310 GTAGATGATGATGATGAAGCATGGAACCTTGGAAGAAAGACACAGAA 261
954 userSerArgSerGlyArgGluSerValSerThrAlaSerAspGlnPro 971
|||||
260 AAGTCTCAAGATCAGGAGAGAGTCTGTATCCACAGCCAGATGATAGCCT 211
971 erHisSerLeuGlnArgGlnMetAsnGlyAsnGlnGluLysGlyAspLys 987
|||||
210 CCCACTCTCGGAGAGACAAATGATGAAACCAAGAAAGGATGATAG 161
988 ThrAspArgLysLysAspLysThrGlyLysGluLysLysLysAspArg 1004
|||||
160 ACTGATAGAAAAAAGATTAACCTGAAAAAGAAAGAAAGAAAGATAGGA 111
1004 pLysGluLysAspLysMetLysAlaLysLysGlyMetLeuLysGlyLeu 1021
|||||
110 TTAGGAGAGAGATTAATGAAGCAAGAGGAAGGATGCTGAGGCGTTGG 61
1021 LysAspMetPheArg 1025
|||||
60 GAGCATGTTCCAGG 47
seq_name: gb_est2:BG910244
seq_documentation_block:
LOCUS BG910244 694 bp mRNA linear EST 05-JUN-2001
DEFINITION 602805735F1 NCI_CGAP_Brn67 Homo sapiens CDNA clone IMAGE:4937865
5' mRNA sequence.
ACCESSION BG910244
VERSION BG910244.1 GI:14290720
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 694)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: David N. Louis, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L14M10872 row: f column: 10  
 High quality sequence stop: 694.

#### FEATURES

source  
 1..694  
     /organism="Homo sapiens"  
     /db\_xref="taxon:9606"  
     /clone="IMAGE:4937865"  
     /clone\_lib="NCI\_CGAP\_Brn67"  
     /tissue\_type="anaplastic oligodendroglioma with 1p/19q  
     loss"  
     /lab\_host="DH10B (T1 phage-resistant)"  
     /note="Organ: Brain; Vector: pCMV-Sport6; Site\_1: NotI;  
     Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
     Average insert size 2.3 kb. Constructed by Life  
     Technologies. Note: this is a NCI\_CGAP Library."  
 BASE COUNT 186 a 189 c 205 g 114 t  
 ORIGIN

alignment\_scores:                      Quality: 179.00                      Length: 193  
   Ratio: 0.932                      Gaps: 1  
   Percent Similarity: 99.462           Percent Identity: 99.482

alignment\_block:  
 US-09-757-781-2 x BG910244 ..

Align seg 1/1 to: BG910244 from: 1 to: 694

```

1108 MetAlaLeuAsnAlaArgProGlnSerProArgGluGlnHisMetMetas 1124
|||||
43 ATGGCTTCAACGCTAGACCTCAGAGCCACAGAAAGGATATGATGGA 92
1124 pAlaLeuTyrlaGlnValLysLysProArgAsnSerIysProSerProy 1141
|||||
93 TGCCTTTGTATGCCCAAGTCAAGAAACCCGCGAATTCCAAACCTTCACCTG 142
1141 alAspSerAsnArgSerThrProSerAsnHisAspArgIleGlnArgLeu 1157
|||||
143 TAGACAGTACACGATCAACTCTTAGCAATCATGATGATACAGGCTCTG 192
1158 ArgGlnGluPheGlnGlnAlaLysGlnAspGluAspValGluAspArg 1174
|||||
193 AGCCAAAGATTTCAGCAAGCAAAAGCAAGATGAAATGTAGAAAGATCGTCG 242
1174 gArgThrTyrserPheGlnGlnProTrpProAsnAlaArgProAlaThr 1191
|||||
243 GCGGACCTATAGTTTGTAGCAACCTTGCCGCAACCCAGCGCGGAGACG 292
1191 InSerGlyArgHisSerValSerValGluValAlaGlnMetGlnArgGln 1207
|||||
293 AGAGCGGGCGGCACTCGGTCTCGTGGAGCTGCAGATGCAGGCGCAAGCG 342
1208 GlnGlnGluArgGluSerSerGlnGlnAlaGlnArgGlnTyrserSer 1224
|||||
343 CAGGAGGAGCGCGAGAGCTCCACAGCAGGCCAGCCAGTACAGCTCTCT 392
1224 uProArgGlnSerArgLysAsnAlaSerSerValSerGlnAspSerTrp 1241
|||||
393 GCCTCGGCAAGCAGGAAATAGCCAGCTCGGTCTCCACAGACACTCTTGGG 442

```

```

1241 lucInasntYrSerProgluglypheGlnSerAlaIysGluAsnPro 1257
|||||
443 AGCAAGAACTACTCCCTGGGAAAGGCTCCAGAGTGCCAAAGAGAAACCCC 492
1258 ArgTYrSerSerTYrGlnIySerArgAsnGlyTYrLeuGlyGlyHISgl 1274
|||||
493 AGGTACTCCAGCTACCAAGGCTCCAGAAAGGCTCCTGGGAGGACATGG 542
1274 yPheAsnAlaArgValMetLeu.GluThrGlnGluLeuLeuArgGlnGlu 1290
|||||
543 CTTCAACGGCAGGCTCATGCTGGGAAACACAGAGCTCCTCGCCAGAA 592
1291 GluArgArgIySGluGlnGlnMetIys 1299
|||||
593 CAGAGGCGGAGAGGACGACAGATGAAG 619

seq_name: gb_est2:B1093826

seq_documentation_block:
LOCUS B1093826 809 bp mRNA linear EST 20-JUN-2001
DEFINITION 60286060P1 NIH_MGC_10 Homo sapiens CDNA clone IMAGE:5001744 5',
mRNA sequence.
ACCESSION B1093826
VERSION B1093826.1 GI:14512156
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 809)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11035 row: 1 column: 01
High quality sequence stop: 759.
Location/Qualifiers
1. 809
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site:1; NotI;
Site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
FEATURES
source
BASE COUNT 252 a 174 c 191 g 192 t
ORIGIN
alignment_scores:
Quality: 164.00 Length: 164
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x B1093826 ..
Align seg 1/1 to: B1093826 from: 1 to: 809
577 ValLeuThrProAspGlyThrArgGluPheLeuThrPheGluValProLe 593
|||||

```

```

57 GWTCTTACACCTGATGCACCGAGGAATTTCTGACATTTGAGTCCCACT 106
593 uAsnAspSerGlySerAlaGlyLeuGlyValSerValIySGlyAsnArgS 610
|||||
107 TAAATGATTCAGATCTGCAGGCTTGGTGTCAGTGTCAAAAGCTAACCGGT 156
610 erLYSGluAsnHISAlaAspLeuGlyIlePheValIySerIleIleAsn 626
|||||
157 CAAAAGAAACACGACGACATTTGGGAATCTTTGTCAACTCATTAATAT 206
627 GlyIyAlaAlaIaSerIyAspGlyArgLeuArgValaAsnAspGlnIleII 643
|||||
207 GGAGGAGCAGCATCTAAAGATGGAAGGCTTCGGTGAATGATCAATGAT 256
643 eAlaValaSnGlyIySerLeuGlyIySThrAsnGlnAspAlaMetG 660
|||||
257 AGCAGTTAAATGGAGAAATCCCTGTTGGCAAGACAAACCAAGATGCATGG 306
660 lUThrLeuArgArgSerMetSerThrGlnGlyAsnIySArgIyMetIle 676
|||||
307 AAACCTTAAGAGGTCTATGCTACTGAAGGCATTAACGAGGAATGATC 356
677 GlnLeuIleValAlaArgArgIleSerIyScyAsnGlnLeuIySerPr 693
|||||
357 CAGCTTATTTGTCAGAGGAGAAATGAAGCAAGTCAATGAGCTGAAGTCAAC 406
693 oGlySerProGlyProGlnLeuProIleGluThrAlaLeuAspAspA 710
|||||
407 TGGAGACCCCTGGACCTGAGACTGCGCATTAACAGCTGTGATGATA 456
710 rGlnUArgArgIleSerHISSerIyTYrSerGlyIleGlnIyLeuAsp 726
|||||
457 GAGAACGAGAAATTTCCATTCCTCFACAGTGGATGTAAGGGCTGTGAT 506
727 GluSerProSerArgAsnAlaIaLeuSerArgIleMetGly 740
|||||
507 GAATGCCCAAGCAGAAATCTGCCCTCACTAGATATATGGA 548

seq_name: gb_est2:B1870563

seq_documentation_block:
LOCUS B1870563 575 bp mRNA linear EST 11-OCT-2001
DEFINITION 60339404P1 NIH_MGC_90 Homo sapiens CDNA clone IMAGE:5404067 5',
mRNA sequence.
ACCESSION B1870563
VERSION B1870563.1 GI:16044236
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 575)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12030 row: 9 column: 12
High quality sequence stop: 575.
Location/Qualifiers
1. 575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
FEATURES
source

```



/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NcH1;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.7 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
BASE COUNT 182 a 139 c 128 g 126 t  
ORIGIN

alignment\_scores:  
Quality: 162.00 Length: 162  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-757-781-2 x B1870563 ..

Align seg 1/1 to: B1870563 from: 1 to: 575

309 G1G1Y1Y1A1A1G1U1H1S1G1U1A1S1N1L1E1U1P1H1E1A1R1G1U1A1S1N1S1P1C1Y1I1E1V1A1 325  
|||||  
12 G1G1G1T1A1A1C1T1G1A1C1A1T1G1A1A1A1C1T1T1T1C1G1T1G1A1G1A1T1G1C1A1T1T1G1T 61  
325 1A1R1G1I1E1A1S1N1S1P1C1Y1A1S1P1E1U1A1R1G1A1N1A1R1G1A1R1G1H1E1G1U1A1A1G1I1N1H 342  
|||||  
62 C1A1G1A1T1T1A1A1G1A1T1G1C1G1A1C1C1T1T1C1G1A1A1T1G1A1C1A1T1T1G1A1C1A1C1A1C1A1C 111  
342 1S1M1E1P1H1E1A1R1G1I1N1A1M1E1T1A1R1G1T1H1P1R1O1L1E1I1E1T1P1H1E1N1S1V1A1I1 358  
|||||  
112 A1T1A1G1T1T1C1C1A1A1G1C1A1T1G1C1A1C1C1A1C1C1A1C1C1A1T1T1G1T1C1C1A1G1T1G1T 161  
359 P1R1O1A1I1A1A1S1N1S1P1C1Y1G1U1A1S1N1L1E1U1S1E1R1G1I1N1S1E1R1G1I1U1Y1S1A1S1 375  
|||||  
162 C1C1G1C1A1G1C1A1A1T1A1A1G1C1A1G1A1T1G1A1C1A1C1A1C1A1C1A1C1A1G1A1G1A1A1 211  
375 N1A1N1T1Y1T1Y1S1E1R1S1E1R1A1R1P1H1E1S1E1R1P1R1O1A1S1P1S1E1R1G1I1T1Y1I1E1A1S1P1A1N1A1 392  
|||||  
212 C1A1A1T1A1C1A1T1T1C1A1A1G1C1C1T1T1A1G1C1C1T1G1A1C1A1G1C1A1C1A1T1A1T1G1A1C1A1A1 261  
392 R1G1S1E1R1V1A1A1S1E1R1A1G1I1Y1L1E1N1H1S1T1H1R1V1A1G1I1N1A1R1G1A1P1R1O1A1R1G1E1U1 408  
|||||  
262 G1G1G1T1G1A1A1C1A1G1T1G1C1A1G1G1T1T1C1A1C1G1G1T1G1C1A1G1A1G1C1A1C1C1G1A1C1G1 311  
409 A1S1N1H1S1P1R1O1G1I1U1G1I1N1I1E1A1S1P1S1E1R1H1S1E1R1A1R1G1E1U1P1R1O1H1S1E1R1A1 425  
|||||  
312 A1A1C1A1C1C1C1G1T1G1A1G1A1T1A1G1A1C1T1C1A1C1A1G1A1C1T1A1C1C1T1A1G1C1C1 361  
425 A1H1S1P1R1O1S1E1R1G1I1Y1S1P1R1O1S1E1R1A1P1R1O1A1S1E1R1A1P1R1O1G1I1A1N1V1 442  
|||||  
362 A1C1A1C1C1C1T1G1G1G1A1A1M1C1C1A1C1A1T1C1C1C1C1A1G1C1C1T1G1C1A1C1T1G1A1T1G1 411  
442 A1P1H1E1S1E1R1T1H1R1V1A1S1E1R1S1E1R1G1I1Y1T1Y1A1S1N1T1H1R1L1Y1S1I1E1G1I1Y1S1 458  
|||||  
412 T1A1T1T1A1C1T1A1G1A1C1T1G1T1A1T1A1C1A1C1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1 461  
459 A1R1G1E1U1A1S1N1I1E1G1I1N1E1U1L1Y1S1G1I1Y1H1G1I1U1G1I1Y1 470  
|||||  
462 A1G1C1T1T1A1T1T1C1A1G1C1T1T1A1G1A1A1A1G1T1A1C1A1G1A1A1G1T1 497

seq\_name: gb\_est2:BG473458

seq\_documentation\_block:

LOCUS BG473458 660 bp mRNA linear EST 21-MAR-2001  
DEFINITION 602516255F1 NIH\_MGC\_16 Homo sapiens cDNA clone IMAGE:4648117 5',  
mRNA sequence.  
ACCESSION BG473458  
VERSION BG473458.1 GI:13405733  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 660)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs@email.nih.gov](mailto:cgabs@email.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at: [image.lnl.gov](http://image.lnl.gov)  
Plate: LNCM425 row: m column: 14  
High quality sequence stop: 656.  
Location/Qualifiers

FEATURES  
source

1..660  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:4648117"  
/clone\_lib="NIH\_MGC\_16"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pOHB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCAAGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."  
BASE COUNT 204 a 135 c 172 g 149 t  
ORIGIN

alignment\_scores:  
Quality: 159.00 Length: 159  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-757-781-2 x BG473458 ..

Align seg 1/1 to: BG473458 from: 1 to: 660

570 L1Y1A1A1S1P1C1Y1A1S1P1E1U1A1R1G1A1N1A1R1G1A1R1G1H1E1G1U1A1A1G1I1N1H1 586  
|||||  
88 A1A1C1A1C1A1G1A1T1A1G1A1C1A1T1T1C1T1T1A1C1A1C1T1G1C1A1C1A1G1A1G1A1A1 137  
586 E1L1E1U1T1H1R1P1H1E1G1I1U1A1P1R1O1L1E1I1E1T1P1H1E1N1S1V1A1I1 603  
|||||  
138 T1C1G1A1C1A1T1T1G1A1G1C1C1A1T1T1A1T1G1A1T1T1C1A1G1A1T1T1G1G1T1G1G1T1G1 187  
603 A1L1S1E1R1V1A1A1S1E1R1G1I1Y1S1P1R1O1S1E1R1S1E1R1A1R1G1E1U1P1R1O1H1S1E1R1A1 619  
|||||  
188 T1C1A1G1T1A1A1A1G1T1A1C1G1T1C1A1A1A1A1G1A1A1C1A1C1A1G1A1T1T1G1G1A1T1C1 237  
620 P1H1E1V1A1I1Y1S1E1R1I1E1A1S1N1G1I1Y1A1I1A1A1S1E1R1Y1S1A1S1P1G1I1Y1A1R1G1E1 636  
|||||  
238 T1T1T1G1C1A1G1C1A1T1T1A1T1A1T1G1A1G1A1C1A1C1A1C1T1T1A1A1G1A1N1T1G1A1A1G1C1T 287  
636 U1A1R1G1V1A1A1S1N1S1P1C1Y1L1E1A1V1A1A1S1N1G1I1Y1G1U1S1E1R1E1U1L1Y1L1 653  
|||||  
288 T1C1G1G1T1G1A1A1T1G1A1C1A1T1G1A1G1A1C1A1T1T1A1T1G1A1A1T1G1G1T1G1G1C1A1 337  
653 Y1S1T1H1A1S1N1G1I1A1S1P1A1M1E1T1U1H1R1E1U1A1R1G1A1R1G1S1E1R1H1S1E1R1H1C1U1 669  
|||||  
338 A1G1A1C1A1A1C1A1A1G1A1T1G1C1A1A1A1C1C1T1A1A1A1A1G1T1A1T1G1T1A1C1T1A1 387  
670 G1Y1A1N1Y1S1A1R1G1I1Y1M1E1T1I1E1A1S1N1G1I1Y1A1I1A1A1S1E1R1Y1S1A1S1P1G1I1E1S1E1R1Y1 686  
|||||  
388 G1C1A1T1A1A1A1C1A1G1A1T1A1T1C1A1C1C1T1T1T1G1T1G1C1A1A1G1A1A1A1T1A1A1G1A1A1 437  
686 S1C1Y1A1S1N1G1I1U1E1U1Y1S1E1R1P1R1O1G1I1Y1S1E1R1P1R1O1G1I1Y1P1R1O1G1I1U1E1U1P1R1O1I 703

```

|||||
438 GTGCATAGCTGAAGTACCTGGGACCCCTGGAGCTGAGTCCCA 487
703 |egluThralaLeuAspArgLuarArgTlEserHisSerLeuYr 719
488 TTGAACAGCCTTGATGATAGAGAACGAAGAAATTCATTCCTCTAC 537
720 SerGlyIleGluLeuAspGluSer 728
538 AGTGGATTTAGGGCTTGATGATCG 564
seq_name: gb_est2:BG283908

```

```

seq_documentation_block:
LOCUS      BG283908              477 bp    mRNA    linear    EST 21-FEB-2001
DEFINITION 602407659F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4519695 5',
ACCESSION  BG283908
VERSION     BG283908.1
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens

```

```

REFERENCE  1
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE      NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL    National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT    Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/DRP
            CDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10416 row: f column: 16
            High quality sequence stop: 477.
            Location/Qualifiers

```

```

FEATURES
source
1..477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4519695"
/clone_id="NIH_MGC_91"
/tissue_type="adipocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/site_2="SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 Kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC library."
BASE COUNT 171 a 86 c 134 g 86 t
ORIGIN

```

```

alignment_scores:
Quality: 158.00      Length: 158
Ratio: 1.000         Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

```

alignment\_block:

US-09-757-781-2 x BG283908 ..

Align seg 1/1 to: BG283908 from: 1 to: 477

```

858 AlaAspGluThrLysLeuAsnThrValAspAspGlnLysAlaGlySerPr 874
|||||
2 GCTGACGAGACTAACTCAATACAGTGCATGACACAGAAAGCAAGTTCTCC 51
874 GSeArgspAlaGlyProSerLeuGlyLeuLysLysSerSerLeuG 891
|||||
52 CAGCAGAGATGTGGTCTTCCTGGGTCTGAAGAGTCAAGCTCATTTGG 101

```

```

891 |usLeuGlnThrAlaValAlaGluValThrLeuAsnGlyAspIlePro 907
|||||
102 AGACTGTGCAGACCGCATGTCGCGAGGTGACTTGAATGGGGATATTCCT 151
908 PheHisArgProArgProArgIleIleArgGlyArgGlyAsnGluSe 924
152 TTTCATGTCGTCACGGCCGCGATATATCAGACGAGGGAGTCAATGAG 201
924 rPheArgAlaAlaIleAspLysSerTyraAspLysProAlaValAsp 941
202 CTTTCAGAGCTGCATCGACAAATTTATGATAAACCCTGGGTAGATG 251
941 sPAspGluGlyMetGluThrLeuGluLuarAspThrGluGluSerArg 957
252 ATGATGAAGGCATGAGACCTTGGAAGAAAGACAGAAAGAACTTCAAG 301
958 SerGlyArgGluSerValSerThrAlaSerAspGlnProSerHisSer 974
302 TCAAGGAGAGAGTCTGTATCCACAGCAGTATCATGCTTCCACTCTCT 351
974 uGluArgGlnMetAsnGlyAsnGlnGluLysGlyAspLysThrAspArg 991
352 GGAAGACCAAAATGAATGGAACCAAGAGAAAGTGAATAGACTGATGAA 401
991 ySLysAspLysThrGlyLysGluLysLysAspArgAspLysGluLys 1007
402 AAAGATRAAACTGCAAAAGAAAGAAAGAAATGATGAGTAAAGAGAG 451
1008 AspLysMetLysAlaLysLysGly 1015
452 GATTAATAAGAAAGCAAGAGGGA 475
seq_name: gb_est2:BG170486

```

```

seq_documentation_block:
LOCUS      BG170486              713 bp    mRNA    linear    EST 06-FEB-2001
DEFINITION 602322859F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4425975 5',
ACCESSION  BG170486
VERSION     BG170486.1
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens

```

```

REFERENCE  1
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE      NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL    National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT    Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: The I.M.A.G.E. Consortium/LLNL
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10172 row: e column: 16
            High quality sequence stop: 710.
            Location/Qualifiers

```

```

1..713
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4425975"
/clone_id="NIH_MGC_89"
/tissue_type="hypernephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/site_2="SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 Kb. Library enriched for
full-length clones and constructed by Life Technologies.

```



```

|||||
377 TGATGATGATGAAGCATGAGACTTTGGAAAGACACAGAGAAAGTT 328
956 eArYrSeRgIyArGluSerValSerThrAlaSerAspGlnProSerHis 972
327 CAAGATCAGGAGAGACTGTATCCACACCCAGTGATCAGCCCTCCAC 278
973 SerLeuGluArGluMetAsnGlnGlnGlnGlnGlnGlnGlnGlnGln 989
277 TCTCTGGAGAGACAATGATGAAACCAAGAAAGGTATAGACTGA 228
989 pArGlyLysAspLysThrGlyLysGlnLysLysLysAspArgLysG 1006
227 TACAAAAAGGATAAACTGGAAAAAGAAAGAAAGATAGATAGAG 178
1006 LuYsAspLysMetLysAlaLysLysGlyMetLeuLysGlyLeuGly 1022
177 AGAAGATTAATAATGAACCAAGAGAGGATGCTGAAGGCGCTTGGAG 128
1023 MetPhe 1024
|||||
127 ATGTTC 122

```

seq\_name: gb\_est2:BG260547

seq\_documentation\_block:

LOCUS BG260547 966 bp mRNA linear EST 13-FEB-2001  
 DEFINITION 60237096F1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4480366 5',  
 mRNA sequence.

ACCESSION BG260547

VERSION BG260547.1 GI:12770363

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 966)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

Clone sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LHAM10313 row: O column: 23

High quality sequence stop: 723.

Location/Qualifiers

1..966

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4480366"

/clone\_id="NIH\_MGC\_93"

/lssue\_type="transitional cell papilloma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: bladder; Vector: pCMV-SPORT6; Site: 1; NotI;

Site: 2; Salt: Cloned unidirectionally; oligo-dt primed.

Average insert size 1.7 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH\_MGC Library."

BASE COUNT 282 a 209 c 271 g 204 t

ORIGIN

alignment\_scores:

Quality: 149.00 Length: 149

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-757-781-2 x BG260547 ..

Align seg 1/1 to: BG260547 from: 1 to: 966

```

592 ProLeuAsnAspSerGlySerAlaGlyLeuGlyValSerValLysGlyAs 608
|||||
1 CCACCTTAATGATTCAGAGATCGAGCCCTTGGTGCAGTGCACAAAGCTAA 50
608 nArGSerLyGluAsnHisAlaAspLeuGlyIlePheValLysSerIleI 625
|||||
51 CCGGTCAAAAAGAGAACACAGCAGATTTGGCAATCTTTGTCAGATCCATTA 100
625 LeAsnGlyAlaAlaSerLysAspGlyArgLeuArgValAsnAspGln 641
|||||
101 TTATATGAGAGAGCAGCATTTAAAGATGGAAGGCTTGGGTGATATATCA 150
642 LeuIleAlaValAsnGlyLysLeuLeuGlyLysThrAsnGlnAspAl 658
|||||
151 CTGATAGCAGTAAATGAGAAATCCCTGTTGGCAAGACAAACCAAGATGC 200
658 aMetGluThrLeuArgArgSerMetSerThrGluGlyAsnLysArgGly 675
|||||
201 CARGAAACCCCTAAGAGGGCTATGCTACTGAAAGCAATTAACGAGAA 250
675 eTlleGlnLeuIleValAlaArgArgIleSerLysCysAsnGluLeuLys 691
|||||
251 TGATTCAGCTTATTTGTTCAGAGAGATAAGCAAGTGCATATGCTGAAG 300
692 SerProGlySerProProGlyProGluLeuProIleGluThrAlaLeuAs 708
|||||
301 TCACCTGGAGAGCCCTGAGCTGAGCTGAGCCCATTTGAACACCGTTGGA 350
708 pAspArgGluArgArgIleSerHisSerLeuThrSerGlyIleGluGlyL 725
|||||
351 TGATAGAGACGAGAGAAATTTCCATTTCCCTTACAGTGGGATTTGAGGG 400
725 euAspGluSerProSerArgAsnAlaAlaLeuSerArgIleMetGly 740
|||||
401 TTGATGATATCGGCCAGCAAAATGCTGCCCTGAGTAGTAGTATGGGT 447

```

seq\_name: gb\_est2:BG170054

seq\_documentation\_block:

LOCUS BG170054 987 bp mRNA linear EST 06-FEB-2001  
 DEFINITION 602321514F1 NIH\_MGC\_89 Homo sapiens cDNA clone IMAGE:4424954 5',  
 mRNA sequence.

ACCESSION BG170054

VERSION BG170054.1 GI:12676757

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 987)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LHAM10169 row: K column: 03

High quality sequence start: 7

High quality sequence stop: 633.

Location/Qualifiers

1..987

/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="IMAGE:4424954"
/clone_lib="NIH_MGC_89"
/tissue_type="hyponephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: kidney; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

BASE COUNT      297 a      234 c      235 g      221 t
ORIGIN

alignment_scores:
    quality: 147.00      length: 147
    ratio: 1.000      gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x BG170054 ..

Align seg 1/1 to: BG170054 from: 1 to: 987

150 GlyleusertHrServalSerAspSerAspHeserSergluGluprose 166
|||||
121 GGCCCTCCACTCTGTCAGTAGATGTAATTTCTCTGAAGAGCCCTC 170
|||||
166 fArlysaAnProthArgrTpsertHrAlaglyPheleuysglna 183
|||||
171 AAGGAAAATCCACACGCTGGTCACACACAGCTGGCTCTCAAGCAGA 220
|||||
183 snrHrAlaglySerProlyThrCysAspArglyLysAspGluAsnTyr 199
|||||
221 ACACGTGCTGGAGTCTTAACACCTGCAGCAGAGAAAGATGAATAACTAC 270
|||||
200 ArgerleuprOargaspPthSerAsnTpsersnGlnPheglnarAs 216
|||||
271 AGAAGCCCTCCGCGGAGATGATGTAACGTCTAACCAATTCAGAGAGA 320
|||||
216 PAnAlarGerSerleusSerAlaSerHISerProMetValglyLysTrpL 233
|||||
321 CAATGCTGCTGCTCTGAGAGTCCAGTCCCAATGTTGGCCAAAGTGGC 370
|||||
233 euglyysgIngluGlnAspGluAspGlyThrGluGluAspAsnSerAr 249
|||||
371 TGGAGAAACAGAACAGATGAGATGGGACAGAAAGAGATTAACAGTCTG 420
|||||
250 ValGluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPh 266
|||||
421 GTTGAACTGTTGGACATGACGACGGGTTGGAGCATATATACCAACTT 470
|||||
266 eSerleuAspMetValLysIleuValGluValProAsnAspGlyLysP 283
|||||
471 TTCTCTGATCATATGTTAAAGCTGCTGTAAGTCCCAACGATGGAGGCG 520
|||||
283 rOlEuglyIleHisValValProPheSerAlaArgGlyGly 296
|||||
521 CTCTGGGAATCATGTAGTGCCTTTCAGTGTGCGAGGGGGA 561
|||||

seq_name: gb_estl:A1990048

seq_documentation_block:
LOCUS      A1990048      442 bp      mRNA      linear      EST 08-MAR-2000
DEFINITION      wsl19e01.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2497656 3',
                mRNA sequence.
ACCESSION      A1990048
VERSION        A1990048.1      GI:5836929
KEYWORDS
SOURCE
ORGANISM      human.
                Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE      1 (bases 1 to 442)
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.dio.lnl.gov/ldbrp/image/image.html
Insert length: 1105      Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 430.
Location/Qualifiers
1. 442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2497656"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/Note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT      88 a      120 c      87 g      147 t
ORIGIN

alignment_scores:
    quality: 145.00      length: 145
    ratio: 1.000      gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x A1990048/rev ..

Align seg 1/1 to reverse of: A1990048 from: 1 to: 442

845 VallystrArGlySerlySerMetAspLeuGlyIleAlaAspGluTh 861
|||||
442 GTTAAACACGAAATCAAAAAGCATGATTTAGGATATGCTGACGAGAC 393
|||||
861 rlyLeuAsnThrValAspAspGlnLysAlaGlySerProSerArgAspY 878
|||||
392 TAAACTCAATACAGTGAACACAGAAAGCAGTCTCCACAGAGATG 343
|||||
878 alGlyProSerleuGlyLeuLysSerSerleuGluSerleuGln 894
|||||
342 TGGGTCTTCCCTGGGCTGGAAGAGTCAAGCTCGTTGAGAGTCTGCG 293
|||||
895 ThrAlaValAlaGluValThrLeuAsnGlyAspIleProPheHisArgP 911
|||||
292 ACCGAGATTGCCGAGGTGACTTTGAATGGGATATTCCTTTCATCGTCC 243
|||||
911 cArgProArgIleIleArgGlyArgGlyCysAsnGluSerPheArgAla 928
|||||
242 ACGGCCGGGATATACAGAGCAGGGGATGCAATGAGAGCTTCAGAGCTG 193
|||||
928 laIleAspLysSerTyraLysLysProAlaValAspAspAspGlyGly 944
|||||
192 CCATGACAAATCTTATGATTAACCCGCGGTAGATGATGATGATGAAGGC 143
|||||

```

```
945 MetGluThrLeuGluGluAspThrGluGluSerArgSerGlyArgI 961
|||||
142 ATGGAGACCTTGGAGAGACACAGAGAAAGTTCAAGATCAGGAGAGA 93
|||||
961 userValserThrAlaSerAspGlnProSerHisSerLeuGluArgGlnM 978
|||||
92 GTCGTATCCACAGCCAGTGCATCCCTCCCTCTCTGGAGAGACAAA 43
|||||
978 etAsnGlyAsnGlnGluLysGlyAspLysThrAsp 989
|||||
42 TGAATGGAACCAAGAGAAAGTGATTAAGACTGAT 8
```

OM of: US-09-757-781-2 to: Issued\_Patents\_NA:\* out\_format: pfs

Date: Jul 24, 2002 6:12 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

#### Command line parameters:

-MODBL-Frame+ .p2n.model -DEV-xlh  
-Q-/cgn2\_1/USPRO.spool/US09757781/runat\_22072002.154317.18613/app\_query.fasta\_1.1434  
-DB-Issued\_Patents\_NA -ORWT-fastap -SUPERfix-olig.rni  
-GAPOP-4.500 -GAPEXT-0.050 -MINMATCH-0.100 -LOOPEXT-0.000  
-LOOPEXT-0.000 -GAPOP-4.500 -GAPEXT-0.050 -XGAPOP-60.000  
-XGAPEXT-60.000 -FEAPOP-6.000 -FEAPEXT-7.000 -YGAPOP-60.000  
-YGAPEXT-60.000 -DELDP-6.000 -DELEXT-7.000 -START-1  
-MATRIX-oligo -TRANS-human40.cdi -LIST-45 -DOCALLIGN-200  
-THR\_SCORE-quality -THR\_MIN-1 -ALIGN-15 -MODE-LOCAL -OUTFMT=pfs  
-NORExt-HEARSize-500 -MINLEN-0 -MAXLEN-200000000  
-USER=US09757781.ecgn1\_1.68 -NCPU-6 -ICPU-3 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -NO\_XLPHY -WAIT -THREADS-1

#### Search information block:

Query: US-09-757-781-2  
Query length: 1356  
Database: Issued\_Patents\_NA\*  
Database sequences: 383533  
Database length: 122816752  
Search time (sec): 65.160000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

```
score_list:
Strd Orig      ZScore  EScore Len  Documentation
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-276-531-77 + 19.00 296.09 8.3e-09 1042
/cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-735-041A-3 + 10.00 137.71 5.53 1812
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-190-476B-3 - 10.00 137.71 5.53 1812
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-190-889A-3 - 10.00 137.71 5.53 1812
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-190-938B-3 - 10.00 137.71 5.53 1812
/cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-735-041A-5 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-735-041A-7 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-190-476B-5 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-190-889A-5 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-190-889A-7 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-190-938B-5 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-190-938B-7 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/lna/5A_COMB.seq:US-07-743-518-23 + 8.00 115.26 98.46 369
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-927-219-42 + 8.00 110.92 171.91 662
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-818-112-20 - 8.00 107.70 259.86 1021
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-818-111-20 - 8.00 107.70 259.86 1021
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-056-556-20 - 8.00 107.70 259.86 1021
/cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-611-757-33 + 8.00 107.40 269.80 1062
/cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-611-757-33 - 8.00 107.40 269.80 1062
/cgn2_6/ptodata/2/lna/PCUTS_COMB.seq:PCUTS-05950-33 + 8.00 107.40 269.80 1062
/cgn2_6/ptodata/2/lna/PCUTS_COMB.seq:PCUTS-05950-33 - 8.00 107.40 269.80 1062
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-928-213B-58 - 8.00 107.08 281.17 1109
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-445-640-5 + 8.00 106.72 294.45 1164
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-170-558-5 + 8.00 106.72 294.45 1164
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-447-314-5 + 8.00 106.72 294.45 1164
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-445-461-5 + 8.00 106.72 294.45 1164
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-004-171-1 - 8.00 106.28 311.56 1235
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-255-368-7 + 8.00 105.94 325.49 1293
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-415-655-14 + 8.00 105.59 340.36 1355
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-936-165A-113 - 8.00 105.10 362.60 1448
/cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-665-037-1 + 8.00 104.16 408.79 1642
/cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-666-067-1 + 8.00 104.16 408.79 1642
/cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-732-870-1 + 8.00 104.16 408.79 1642
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-466-954A-1 + 8.00 100.87 623.63 2557
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-286-305A-2 + 8.00 100.35 666.58 2742
/cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-441-104A-2 + 8.00 100.35 666.58 2742
/cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-440-816A-2 + 8.00 100.35 666.58 2742
```

```
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-417-381A-2 + 8.00 100.35 666.58 2742
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-965-762-21 - 8.00 98.88 805.24 3343
/cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-231-227-3 + 8.00 98.86 807.31 3352
/cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-445-640-1 + 8.00 98.30 866.69 3611
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-170-558-1 + 8.00 98.30 866.69 3611
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-447-314-1 + 8.00 98.30 866.69 3611
/cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-445-461-1 + 8.00 98.30 866.69 3611
```

seq\_name: /cgn2\_6/ptodata/2/lna/6B\_COMB.seq:US-09-276-531-77

seq\_documentation\_block:

Sequence 77, Application US/09276531

Patent No. 6183968

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Lal, Preeti

APPLICANT: Hillman, Jennifer L.

APPLICANT: Yue, Henry

APPLICANT: Reddy, Roopa

APPLICANT: Guegler, Karl J.

APPLICANT: Baughn, Mariah R.

TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING

RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION.

NUMBER OF SEQUENCES: 134

CORRESPONDENCE ADDRESSES:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/276,531

FILING DATE: Herewith

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/079,677

FILING DATE: March 27, 1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Lynn E. Murry, Ph.D.

REGISTRATION NUMBER: 42,918

REFERENCE/DOCKET NUMBER: PA-0008 US

TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 77:

SEQUENCE CHARACTERISTICS:

LENGTH: 1042 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: BLADTUT04

CLONE: 1555118

US-09-276-531-77

alignment\_scores:

Quality: 19.00

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-09-757-781-2 x US-09-276-531-77

US-09-757-781-2 ..

Align seg 1/1 to: US-09-276-531-77 from: 1 to: 1042

882 LeuglyleuLysSerSerSerleuglnThrAlaVal 898  
|||||  
662 CTGGCTTGAAAAAGTCCAGCTCTGGAGAGTCTGCACACTGCAGTGGC 711  
898 aglval 900  
|||||  
712 CGAGGTC 718

seq\_name: /cgn2\_6/ptodata/2/lna/5b\_COMB.seq:US-08-735-041A-3

seq\_documentation\_block:

; Sequence 3, Application US/08735041A

; Patent No. 5914251

; GENERAL INFORMATION:

; APPLICANT: Farrell, Catherine L.

; APPLICANT: Martin, Francis H.

; APPLICANT: Yabkowitz, Rachel

; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH

; TITLE OF INVENTION: FACTOR

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Amgen Inc.

; STREET: 1840 De Havilland Drive

; CITY: Thousand Oaks

; STATE: California

; COUNTRY: USA

; ZIP: 91320-1789

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/735,041A

; FILING DATE: 22-OCT-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Mazza, Richard J.

; REGISTRATION NUMBER: 27,657

; REFERENCE/DOCKET NUMBER: A-414

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 805 447 4112

; TELEFAX: 805 499 6751

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1812 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; US-08-735-041A-3

alignment\_scores:

Quality: 10.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-09-757-781-2 x US-08-735-041A-3/rev ..

Align seg 1/1 to reverse of: US-08-735-041A-3 from: 1 to: 1812

420 ArgleuproHisSerAlaHisProSerGly 429

|||||

1681 CGGCTGCCACATTCGCACATCCCTCCGGG 1652

seq\_name: /cgn2\_6/ptodata/2/lna/6a\_COMB.seq:US-09-190-476B-3

seq\_documentation\_block:

; Sequence 3, Application US/09190476B

; Patent No. 6025204

; GENERAL INFORMATION:

; APPLICANT: Farrell, Catherine L.

; APPLICANT: Yabkowitz, Rachel

; APPLICANT: Martin, Francis H.

; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH

; TITLE OF INVENTION: FACTOR

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Amgen Inc.

; STREET: 1840 De Havilland Drive

; CITY: Thousand Oaks

; STATE: California

; COUNTRY: USA

; ZIP: 91320-1789

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/190,476B

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/735,041

; FILING DATE: 22-OCT-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Mazza, Richard J.

; REGISTRATION NUMBER: 27,657

; REFERENCE/DOCKET NUMBER: A-414

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 805 447 4112

; TELEFAX: 805 499 6751

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1812 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; US-09-190-476B-3

alignment\_scores:

Quality: 10.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-09-757-781-2 x US-09-190-476B-3/rev ..

Align seg 1/1 to reverse of: US-09-190-476B-3 from: 1 to: 1812

420 ArgleuproHisSerAlaHisProSerGly 429

|||||

1681 CGGCTGCCACATTCGCACATCCCTCCGGG 1652

seq\_name: /cgn2\_6/ptodata/2/lna/6a\_COMB.seq:US-09-190-889A-3

seq\_documentation\_block:

; Sequence 3, Application US/09190889A

; Patent No. 6075008

; GENERAL INFORMATION:

; APPLICANT: Farrell, Catherine L.

; APPLICANT: Martin, Francis H.

; APPLICANT: Yabkowitz, Rachel

; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH

; TITLE OF INVENTION: FACTOR

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Amgen Inc.

; STREET: 1840 De Havilland Drive

; CITY: Thousand Oaks

; STATE: California



```

; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/190,889A
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/735,041
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/DOCKET NUMBER: A-414
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 805.447.4112
; TELEFAX: 805.499.6751
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1812 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-09-190-889A-3

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-757-781-2 x US-09-190-889A-3/rev ..
  Align seg 1/1 to reverse of: US-09-190-889A-3 from: 1 to: 1812

420 ArgleuprohisSerAlaHisProSerGly 429
|||||
1681 CGGCTGCCACATTCGACATCCCTCCGGG 1652

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-190-938B-3

seq_documentation_block:
; Sequence 3, Application US/09190938B
; Patent No. 6197939
; GENERAL INFORMATION:
; APPLICANT: Farrell, Catherine L.
; Martin, Francis H.
; Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/190,938B
; FILING DATE: 12-No. 6197939-1998
; CLASSIFICATION: <unknown>
```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/DOCKET NUMBER: A-414
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 805.447.4112
; TELEFAX: 805.499.6751
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1812 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
; US-09-190-938B-3

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-757-781-2 x US-09-190-938B-3/rev ..
  Align seg 1/1 to reverse of: US-09-190-938B-3 from: 1 to: 1812

420 ArgleuprohisSerAlaHisProSerGly 429
|||||
1681 CGGCTGCCACATTCGACATCCCTCCGGG 1652

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-735-041A-5

seq_documentation_block:
; Sequence 5, Application US/08735041A
; Patent No. 5914251
; GENERAL INFORMATION:
; APPLICANT: Farrell, Catherine L.
; Martin, Francis H.
; Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/735,041A
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/DOCKET NUMBER: A-414
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 805.447.4112
; TELEFAX: 805.499.6751
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

MOLECULE TYPE: CDNA  
US-08-735-041A-5

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-757-781-2 x US-08-735-041A-5/rev ..

Align seg 1/1 to reverse of: US-08-735-041A-5 from: 1 to: 2103

420 ArgleuproHisSerAlaHisProSerGly 429  
|||||  
1681 CGGCTGCCACATCTGCACATCCCTCCGG 1652

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-735-041A-7

seq\_documentation\_block:  
; Sequence 7, Application US/08735041A  
; Patent No. 5914251  
; GENERAL INFORMATION:  
; APPLICANT: Farrell, Catherine L.  
; APPLICANT: Martin, Francis H.  
; APPLICANT: Yabkowitz, Rachel  
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH  
; TITLE OF INVENTION: FACTOR  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/735, 041A  
; FILING DATE: 22-OCT-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mazza, Richard J.  
; REGISTRATION NUMBER: 27,657  
; REFERENCE/DOCKET NUMBER: A-414  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 805.447.4112  
; TELEFAX: 805.499.6751  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2103 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; US-08-735-041A-7

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-757-781-2 x US-08-735-041A-7/rev ..  
Align seg 1/1 to reverse of: US-08-735-041A-7 from: 1 to: 2103

420 ArgleuproHisSerAlaHisProSerGly 429  
|||||  
1681 CGGCTGCCACATCTGCACATCCCTCCGG 1652

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-09-190-476B-5

seq\_documentation\_block:  
; Sequence 5, Application US/09190476B  
; Patent No. 6025204  
; GENERAL INFORMATION:  
; APPLICANT: Farrell, Catherine L.  
; APPLICANT: Martin, Francis H.  
; APPLICANT: Yabkowitz, Rachel  
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH  
; TITLE OF INVENTION: FACTOR  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 De Havilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/190,476B  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/735,041  
; FILING DATE: 22-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mazza, Richard J.  
; REGISTRATION NUMBER: 27,657  
; REFERENCE/DOCKET NUMBER: A-414  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 805.447.4112  
; TELEFAX: 805.499.6751  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2103 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; US-09-190-476B-5

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-757-781-2 x US-09-190-476B-5/rev ..

Align seg 1/1 to reverse of: US-09-190-476B-5 from: 1 to: 2103

420 ArgleuproHisSerAlaHisProSerGly 429  
|||||  
1681 CGGCTGCCACATCTGCACATCCCTCCGG 1652

seq\_name: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-09-190-476B-7

seq\_documentation\_block:  
; Sequence 7, Application US/09190476B  
; Patent No. 6025204  
; GENERAL INFORMATION:  
; APPLICANT: Farrell, Catherine L.

```

; APPLICANT: Martlin, Francis H.
; APPLICANT: Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; TITLE OF INVENTION: FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/190,476B
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/735,041
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/DOCKET NUMBER: A-414
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 805.447.4112
; TELEFAX: 805.499.6751
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-190-476B-7

alignment_scores:
      Quality: 10.00      Length: 10
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x US-09-190-476B-7/rev ..
Align seg 1/1 to reverse of: US-09-190-476B-7 from: 1 to: 2103

420 ArgLeuprOHISserAlaHISProSerGly 429
|||||
1681 CGGCTGCCACATTCGCACATCCCTCCGGG 1652

seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq:US-09-190-889A-5

seq documentation_block:
; Sequence 5, Application US/09190889A
; Patent No. 6075008
; GENERAL INFORMATION:
; APPLICANT: Farrell, Catherine L.
; APPLICANT: Martlin, Francis H.
; APPLICANT: Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; TITLE OF INVENTION: FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
```

```

; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/190,889A
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/735,041
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/DOCKET NUMBER: A-414
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 805.447.4112
; TELEFAX: 805.499.6751
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-190-889A-5

alignment_scores:
      Quality: 10.00      Length: 10
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x US-09-190-889A-5/rev ..
Align seg 1/1 to reverse of: US-09-190-889A-5 from: 1 to: 2103

420 ArgLeuprOHISserAlaHISProSerGly 429
|||||
1681 CGGCTGCCACATTCGCACATCCCTCCGGG 1652

seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq:US-09-190-889A-7

seq documentation_block:
; Sequence 7, Application US/09190889A
; Patent No. 6075008
; GENERAL INFORMATION:
; APPLICANT: Farrell, Catherine L.
; APPLICANT: Martlin, Francis H.
; APPLICANT: Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; TITLE OF INVENTION: FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/190,889A
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
```

APPLICATION NUMBER: US 08/735,041  
FILING DATE: 22-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mazza, Richard J.  
REGISTRATION NUMBER: 27,657  
REFERENCE/DOCKET NUMBER: A-414  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 805 447,4112  
TELEFAX: 805 499,6751  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2103 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-09-190-889A-7

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-757-781-2 x US-09-190-889A-7/rev ..

Align seg 1/1 to reverse of: US-09-190-889A-7 from: 1 to: 2103

420 ArgLeuProHISerAlaHisProSerGly 429  
|||||  
1681 CGGCTGCCACATTCGCACATCCCTCCGG 1652

seq\_name: /cgn2\_6/plodata/2/ina/6B\_COMB.seq:US-09-190-938B-5

seq\_documentation\_block:  
Sequence 5, Application US/09190938B  
Patent No. 6197939  
GENERAL INFORMATION:  
APPLICANT: Farrell, Catherine L.  
Martin, Francis H.  
Yabkowitz, Rachel  
TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH  
FACTOR  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 De Havilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/190,938B  
FILING DATE: 12-NO. 6197939-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Mazza, Richard J.  
REGISTRATION NUMBER: 27,657  
REFERENCE/DOCKET NUMBER: A-414  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 805 447,4112  
TELEFAX: 805 499,6751  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2103 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-190-938B-5

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-757-781-2 x US-09-190-938B-5/rev ..

Align seg 1/1 to reverse of: US-09-190-938B-5 from: 1 to: 2103

420 ArgLeuProHISerAlaHisProSerGly 429  
|||||  
1681 CGGCTGCCACATTCGCACATCCCTCCGG 1652

seq\_name: /cgn2\_6/plodata/2/ina/6B\_COMB.seq:US-09-190-938B-7

seq\_documentation\_block:  
Sequence 7, Application US/09190938B  
Patent No. 6197939  
GENERAL INFORMATION:  
APPLICANT: Farrell, Catherine L.  
Martin, Francis H.  
Yabkowitz, Rachel  
TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH  
FACTOR  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 De Havilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/190,938B  
FILING DATE: 12-NO. 6197939-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Mazza, Richard J.  
REGISTRATION NUMBER: 27,657  
REFERENCE/DOCKET NUMBER: A-414  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 805 447,4112  
TELEFAX: 805 499,6751  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2103 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-09-190-938B-7

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-757-781-2 x US-09-190-938B-7/rev ..

Align seg 1/1 to reverse of: US-09-190-938B-7 from: 1 to: 2103

420 ArgLeuProHisSerAlaHisProSerGly 429  
|||||  
1681 CGGCTGCCACATTTTCACATCCCTCCGGG 1652

seq\_name: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:US-07-743-518-23

seq\_documentation\_block:

; Sequence 23, Application US/07743518  
; Patent No. 5397696  
; GENERAL INFORMATION:  
; APPLICANT: YAMAGIHARA, RICHARD  
; APPLICANT: NERURKAR, VIVEK R.  
; APPLICANT: JENKINS, CAROL  
; APPLICANT: MILLER, MARK  
; APPLICANT: GARRUTO, RALPH M.  
; TITLE OF INVENTION: PAPUA NEW GUINEA HUMAN T-LYMPHOTROPIC  
; TITLE OF INVENTION: VIRUS  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1615 L STREET, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20036-5601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,518  
; FILING DATE: 19910812  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26,581  
; REFERENCE/DOCKET NUMBER: WTS/5683/84699/SAP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-861-3000  
; TELEFAX: 202-822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 369 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20..343  
; US-07-743-518-23

alignment\_scores:

Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-757-781-2 x US-07-743-518-23 ..

Align seg 1/1 to: US-07-743-518-23 from: 1 to: 369

766 LeuProValLeuProHisLeu 773  
|||||  
7 CTCGACGTTTCGCCCTCATCTT 30

seq\_name: /cgn2\_6/ptodata/2/lna/6B.COMB.seq:US-08-927-219-42

seq\_documentation\_block:

; Sequence 42, Application US/08927219  
; Patent No. 6187533  
; GENERAL INFORMATION:  
; APPLICANT: Bell, Graeme I.  
; APPLICANT: Yamagata, Kazuya  
; APPLICANT: Oda, Naohisha  
; APPLICANT: Katsaki, Pamela J.  
; APPLICANT: Furuta, Hiroto  
; APPLICANT: Horikawa, Yukio  
; APPLICANT: Menzel, Stephen  
; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY  
; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA  
; NUMBER OF SEQUENCES: 147  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/927,219  
; FILING DATE: Concurrently Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/029,679  
; FILING DATE: 30-OCT-1996  
; PRIOR APPLICATION NUMBER: US 60/028,056  
; FILING DATE: 02-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/025,719  
; FILING DATE: 10-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wilson, Mark B.  
; REGISTRATION NUMBER: 37,259  
; REFERENCE/DOCKET NUMBER: ARCD:272  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 512/418-3000  
; TELEFAX: 512/474-7577  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 662 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 84..188  
; US-08-927-219-42

alignment\_scores:

Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-757-781-2 x US-08-927-219-42 ..

Align seg 1/1 to: US-08-927-219-42 from: 1 to: 662

767 ProValLeuProHisLeuSer 774  
|||||  
46 CCGTGTCCCTCCCTCACCTCTCT 69

Thu Jul 25 08:38:22 2002

us-09-757-781-2.olg.rni

OM of: US-09-757-781-2 to: N.Geneseq\_032802:\* out\_format : pfs  
Date: Jul 24, 2002 6:17 PM  
About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh  
-O=/cgn2.1/USPRO\_pool/US09757781/runat\_22072002\_154318.18636/app-query.fasta.1.1434  
-DB=N.Geneseq\_032802 -OEMT=fastap -SUFFIX=olig.rng -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdl -LIST=45 -DOCALLIGN=200 -THR\_SCORE=quality  
-THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTEMT=pfs -NORM=ext  
-HEADSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09757781@cgn1\_1.238 -NCP=6 -ICP=3 -LONGLOG  
-DEV=TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-757-781-2  
Query length: 1356  
Database: N.Geneseq\_032802:\*  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 322.900000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

## score\_list:

Sequence	Strd Orig	ZScore	EScore Len	! Documentation	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH73337 +		500.00	8713.44	0.0	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH18687 +		420.00	7317.63	0.0	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH78369 +		420.00	7317.63	0.0	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH7293 +		310.00	5394.26	4.1e-2	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH7293 +		195.00	3386.06	2.9e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH33594 +		195.00	3386.06	2.9e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH7883 +		131.00	2270.12	4.2e-1	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAH01182 +		110.00	1889.13	7.1e-7	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH68903 +		92.00	1579.10	1.3e-7	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH71480 +		89.00	1532.08	5.5e-7	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH66057 +		80.00	1447.46	2.8e-7	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH68901 +		78.00	1340.17	2.7e-6	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH23370 +		78.00	1340.17	2.7e-6	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH07109 +		64.00	1094.78	1.2e-5	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH74458 +		60.00	1033.00	3.4e-4	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH54524 +		60.00	1033.00	3.4e-4	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH61963 +		60.00	1024.73	9.9e-4	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH41889 +		55.00	1024.73	9.9e-4	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH61506 +		52.00	937.87	6.8e-4	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH29229 +		55.00	937.87	6.8e-4	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH09806 +		55.00	937.87	6.8e-4	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH35700 +		55.00	937.87	6.8e-4	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH17106 +		55.00	937.87	6.8e-4	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH14145 +		55.00	937.87	6.8e-4	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH68902 +		47.00	797.70	4.4e-3	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH12111 +		47.00	796.96	4.4e-3	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH71413 +		44.00	748.95	2.3e-3	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH71048 +		44.00	748.95	2.3e-3	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH69243 +		42.00	713.15	2.2e-3	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAH2894 +		38.00	647.76	9.8e-2	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAH56005 +		37.00	631.84	7.6e-2	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH70055 +		34.00	570.12	2.1e-2	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH60546 +		19.00	312.81	4.5e-0	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH60570 +		19.00	312.81	4.5e-0	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH60283 +		19.00	312.72	4.5e-0	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH60284 +		19.00	312.69	4.5e-0	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH60347 +		19.00	312.28	4.8e-0	

/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH80567 +	19.00	302.08	1.8
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH0861 +	19.00	296.91	3.4
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH60956 +	19.00	296.91	3.4
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH60963 +	19.00	296.91	3.4
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH61007 +	19.00	296.91	3.4
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH60733 +	15.00	242.76	3.6
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH67317 +	11.00	167.03	0.5

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH73337

seq\_documentation\_block:

ID AAH73337 standard; cDNA; 5510 BP.

AAH73337;

19-SEP-2001 (first entry)

Human cervical cancer marker nucleic acid 4611.

Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

Homo sapiens.

W0200142467-A2.

14-JUN-2001.

08-DEC-2000; 2000WC-US33312.

08-DEC-1999; 99US-0169681.

21-DEC-1999; 99US-0171350.

14-MAR-2000; 2000US-0189315.

12-MAY-2000; 2000US-0203791.

09-JUN-2000; 2000US-0210600.

21-JUN-2000; 2000US-0220114.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Deeds J, Berger A, Zhao X;

WPI; 2001-375006/39.

New isolated nucleic acid for diagnosing and treating cervical cancer and for assessing and detecting compounds for treating the cancer -

Claim 1; Page 1029-1031; 1051pp; English.

The invention relates to novel genes (AAH68727-AAH73383) associated with

cervical cancer with cytostatic activity. The nucleic acids and encoded

polypeptides are useful: to assess if a patient is afflicted with

cervical cancer or has a pre-malignant condition; to monitor the

progression of cervical cancer or a premalignant condition in a patient;

and to select and/or assess the efficacy of a compound or therapy for

inhibiting cervical cancer in a patient. The nucleic acids may also be

useful for gene therapy.

Sequence 5510 BP: 1526 A; 1242 C; 1418 G; 1305 T; 19 other;

alignment\_scores:

Quality: 500.00 Length: 500

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-757-781-2 x AAH73337 ..

Align seg 1/1 to: AAH73337 from: 1 to: 5510

857 ileaiaapglutThrLysleuasnThrValaiaapaglnLyslaag1se 873

|||||

2299 ATAGCTGACGACACCTAAACTCATATACATGATGACACCAAGACAGGTTTC 2348





PR	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	
DR	WPI: 2001-318749/34.	
XX		
XX	Primer sets for synthesizing polynucleotides, particularly the 5602	
PT	full-length cDNAs defined in the specification, and for the detection	
PT	and/or diagnosis of the abnormality of the proteins encoded by the	
PT	full-length cDNAs -	
XX		
PS	Claim 8; SEQ ID 18774; 2537pp + CD ROM; English.	
XX		
CC	The present invention describes primer sets for synthesizing 5602	
CC	full-length cDNAs defined in the specification. Where a primer set	
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary	
CC	to the complementary strand of a polynucleotide which comprises one of	
CC	the 5602 nucleotide sequences defined in the specification, where the	
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination	
CC	of an oligonucleotide comprising a sequence complementary to the	
CC	complementary strand of a polynucleotide which comprises a 5'-end	
CC	sequence and an oligonucleotide comprising a sequence complementary to a	
CC	polynucleotide which comprises a 3'-end sequence, where the	
CC	oligonucleotide comprises at least 15 nucleotides and the combination of	
CC	the 5'-end sequence/3'-end sequence is selected from those defined in	
CC	the specification. The primer sets can be used in antisense therapy and	
CC	in gene therapy. The primers are useful for synthesizing polynucleotides	
CC	particularly full-length cDNAs. The primers are also useful for the	
CC	detection and/or diagnosis of the abnormality of the proteins encoded by	
CC	the full-length cDNAs. The primers allow obtaining of the full-length	
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and	
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to	
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632	
CC	represent oligonucleotides, all of which are used in the exemplification	
CC	of the present invention.	
XX		
SO	Sequence 2718 BP; 840 A; 586 C; 688 G; 604 T; 0 other:	
	alignment_scores:	
	Quality: 420.00 Length: 508	
	Ratio: 0.830 Gaps: 1	
	Percent Similarity: 99.606 Percent Identity: 99.606	
	alignment_block:	
	US-09-757-781-2 x AAH18587 ..	
	Align seg 1/1 to: AAH18587 from: 1 to: 2718	
234	GIUYSGLNGUGLNASPGIUASPGIYThrGluGUASPAASerArgVA 250	
327	GAGAAACACAGAACAGAGATGGGACACAAAGGATTAACAGTCGCGT 376	
250	IGIUProValGlyNHISAlAaPrThGlyLeuGluNHISIlleProAsnPh 267	
377	TCAACTGTGGACATGCTGTGACACGGGTTTGAGACATATACCAACTTTT 426	
267	erLeuAsPAspMetValLysLeuValGluValProAsnAspGlyGlyPro 283	
427	CTCTGGATGATATGGTTAAAGCTCGTAGAAGTCCCAACAGATGGAGGGCT 476	
284	LeuGlyIleHisValAlaProPheserAlaArgGlyIleArgThrLeuGI 300	
477	CTGGGAATCCATGTAGTCCTTTCAGTCTCTGAGCGCGCAACACCTGGG 526	
300	YeuLeuValLysArgLeuGluIuysGlyGlyAlaGluHisGluAsnL 317	
527	GTTATTTAGTAAACGATTGGAGAAAGGTGTAAAGCTGTAACATGTAATC 576	
317	eupheArGluAsnAspCysIleValAlaArgIleAsnAspGlyAspLeuArg 333	
577	TTTTTCGTGAGAAATGATTCATGTGCAGAGATTAAATGATGGGACCTTGA 626	
334	AsnArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgTh 350	

```

650 euleuglylstrhrasnglnaspalametglutthreuarargsermet 666
|||||
1576 TGTGGGCAAGACAAACCAAGATGCATGAAACCTTAAGAAAGTCTATG 1625
667 SerThgluylvsnlysaraglymetllglnleuilevalalalargar 683
|||||
1626 TCACACGAAGGCATATACAGAGAAATGATCCAGCTTATGTTGCAAGGAG 1675
683 gllserlyscysasngluleuylserprogllyserprogllyprog 700
|||||
1676 AATACCAAGTCAATGACCTGAACCTGAGAGCCCTGACCTG 1725
700 luleuprolleugluthralaleuaspararglualargylleserhis 716
|||||
1726 AGCTGCCCTTCAACACGCTGTGATGATGAGAACCAAGATTTCCCAT 1775
717 Serleuylrserglylleglulyleuaspclyserprogllyarganal 733
|||||
1776 TCCTCTACAGTGGGATTGAGGGGCTTGATGAAATGCCAGCAAAATGC 1825
733 AlaleuSerArglleMetgly 740
|||||
1826 TCCTCTCAGTAGATTAATGGGT 1847

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH78369
seq_documentation_block:
ID AAH78369 standard; DNA; 2718 BP.
AC
AAH78369;
DT 26-NOV-2001 (first entry)
XX
XX Nucleotide sequence of a human protein kinase/protein phosphatase.
XX
XX Human; protein kinase; protein phosphatase; signal transduction; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 33..2630
XX /*tag= a
XX /product= "protein kinase/protein phosphatase"
XX
XX MO200109316-A1.
XX
XX 08-FEB-2001.
XX
XX 28-JUL-2000; 2000MO-IP05061.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 18-OCT-1999; 99US-0159590.
XX 11-JAN-2000; 2000JP-0118776.
XX 17-FEB-2000; 2000US-0183322.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;
XX Senoo C, Nezu J;
XX
XX MPI: 2001-570286/64.
XX P-PSDB; AAG67610.
XX
XX New genes encoding proteins with protein kinase/protein phosphatase
XX activity, useful in the diagnosis and treatment of diseases -
XX
XX Claim 1, Page 67-76; 233bp; Japanese.
XX
XX The present sequence encodes a human protein kinase/protein phosphatase.

```

```

CC It is expected that the protein kinase/protein phosphatase gene
CC participates in signal transduction in cells. The protein kinase/protein
CC phosphatase polypeptides and polynucleotides are useful for developing
CC diagnostics and treatment agents for human and animal diseases. The
CC protein kinase/protein phosphatase polypeptides are useful as target
CC molecules in designing novel drugs. The protein kinase/protein
CC phosphatase polynucleotides are useful as a source of probes and
CC primers, which may be used to isolate homologous sequences.
XX
XX Sequence 2718 BP; 840 A; 586 C; 688 G; 604 T; 0 other:

alignment_scores:
Quality: 420.00 Length: 508
Ratio: 0.830 Gaps: 1
Percent Similarity: 99.606 Percent Identity: 99.606

alignment_block:
US-09-757-781-2 x AAH78369 ..

Align seq 1/1 to: AAH78369 from: 1 to: 2718

234 GLUUGSLUGLUGLNASPGLUASPGLYTHRGLUGLUSPASNSERARGVA 250
|||||
327 GAGAACCAAGACAGATGAGATGGGACAGAAAGATTAACAGTCGTCT 376
250 LGUUPROVALGLYHISALASPTHRGLYLEUGLUNHISLEPROASNP 267
|||||
377 TGAACCTGTTGGACATGCTGACACGCGTTTGGAGCATATACCACTTT 426
267 ERLEUASPMETVALYLSLEUVALGLUVALPROASNPGLYGLYPRO 283
|||||
427 CTCTGATATATGTTAAAGCTCGTAGAAATGCCCAAGATGAGGGCT 476
284 LEUGLYLLEHISVALVALPROPHESERALARGLYGLYARGTHRLEUG 300
|||||
477 CTGGGAATCCATGTAGTCCCTTCACTGCTCGAGCGCGAAGACCTCGG 526
300 YLEUVALYLSARGLEUGLULYSGLYGLYVALAGLUNHISGLUASNL 317
|||||
527 GTTATTATTAAACGATTTGGAGAAAGGTGGTAAAGCTGAACATGAAATC 576
317 EUPHEATRGULNASNPCTSLLEVALRGLILEASNPGLYASPLEUATG 333
|||||
577 TTTTCGTGAGAAATGATGTCATGTCAGATTAATGATGACCACTTGA 626
334 ASNARGARPHGGLUGLINALAGLHISMETPHEATRGINALAMETARG 350
|||||
627 AATGAGAGATTTGACACAGCAACATATGTTCCGCAAGCCATCGTAC 676
350 RPOLLIELETRPHEHISVALVALPROALALASNPYSGLUNHIS 367
|||||
677 ACCCATCATTTGGTTCATGTGGTCTCTCAGCAAAATTAAGACGATG 726
367 LUGLLEUSERGINSERGLULYSASNANTYRYSERSEARGPHESER 383
|||||
727 AACAACTATCCCAAGTGAAGAAACAATTACTATTCAAGCGTTTACG 776
384 PROASPSERGINTYRILEASPNARGSERVALANSERIALAGLYLEUN 400
|||||
777 CCGACAGCCAGCATATATGACACAGAGATGACAGTGAAGTGGAGGGCTTCA 826
400 STRVALIGINARGALAPROARGLEUASNHISPROGLUGLINTLEASP 417
|||||
827 CACGGTGCAGAGACCCCGACTGAACCCCGCTGAGAGATGACT 876
417 ERHISERARGLEUPROHISERIALHISPROSERGLYLSPROPROSER 433
|||||
877 CTCACCTCAAGACTACTCATGACCAACCCCTCGGAAAAACCATCC 926
434 ALAPROALASERIALAPROGLNASNPVALPHESERTHRVALSERSE 450
|||||
927 GCTCAGCCTCGGACCTCAGAAATGATTTAGACGACTGTAAAGCTGG 976

```

```

450 YTYRAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysG 467
|||||
977 TTYTAAACACCAAAAAATAGCAAGAGCTTAATATCCAGCTTAAGAAAG 1026
|||||
467 LyrThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 483
|||||
1027 GTTACAGAAAGTTTGGGATCTCAGATCACTTCCACAGATGTAAACAATAGGT 1076
|||||
484 GlySerAlaProIleTyrValLysAsnIleLeuProArgIleAlaAlaI 500
|||||
1077 GGCTCAGACTCCAAATCTATGTGAAAAACATTCTCCCGGGGGGGGGCAT 1126
|||||
500 eGlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGly 517
|||||
1127 TCAGAGATGGCCGACTTAAGCAGAGACAGACTTATAGAGCTAAATGAG 1176
|||||
517 AlaAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSer 533
|||||
1177 TAGATTATAGTGGCAATCCCAAGAGAAAGTTGTTGCTGCTTGAGAAAGC 1226
|||||
534 ThrLysMetGluGlyThrValSerLeuLeuValPheArgGlnIleAspAl 550
|||||
1227 ACCAAGATGGAAGAACTGTGAGCCTCTGTGCTTGGCCAGAAAGACG 1276
|||||
550 aPheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProL 567
|||||
1277 CTTCACACCCAAAGGAACTGAATCAGAGCCCAAGCCAGATGCAGATTCCA 1326
|||||
567 ysgGluThrLysAlaGluAspGlnAspIleValIleuThrProAspGlyThr 583
|||||
1327 AAGAAACGAAAGCAGAAAGATGAGATGTGTTCTTACACCGATGGACCC 1376
|||||
584 ArgGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaG 600
|||||
1377 AGGAAATTTCTGACATTGGAAGTCCCACTTAGT.GATTCAAGATCTGAG 1425
|||||
600 LysGluGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAsp 616
|||||
1426 GCCTTGGTGTGCAGTGCAAAGGTAAACCGGTCAAAAGAGAAACACGCA 1475
|||||
617 LeuGlyIlePheValLysSerIleIleAsnGlyGlyAlaIleSerLysAs 633
|||||
1476 TTGGGAATCTTTGTCAAGTCCATTATTAATGAGAGAGCAGATCTTAAGA 1525
|||||
633 pGlyArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerL 650
|||||
1526 TGGAAAGCTTCGGGTGATGATCAACTGATGACGTAATGAGAAATCCC 1575
|||||
650 euleuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgSerMet 666
|||||
1576 TGTGTGGCAAGACAAACCAAGATCCATGGAACCTTAAGAGGTCTATG 1625
|||||
667 SerThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaIleArg 683
|||||
1626 TCTTCTGAAGCATTAAACGAGAAATGATCCAGCTTATGTGCAAGAGAG 1675
|||||
683 gIleSerLysCysAsnGluLeuLysSerProGlySerProGlyProG 700
|||||
1676 AATTAAGCAAGTGAATGAGTGAAGTCACTGGGAGCCCCCTGAGACTG 1725
|||||
700 IuleuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHis 716
|||||
1726 AGCGGCCCATTTGAAACAGCGTTGGATGATAGAACAGAAATTTCCCAT 1775
|||||
717 SerLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAl 733
|||||
1776 TCCCTCTACAGATGGGATTTGAGGGGCTTGATGAATCGCCACAGAAATGC 1825
|||||
733 AlaLeuSerArgIleMetGly 740
|||||
1826 TGCCCTCAGTGAATGAATGGGT 1847

```

```

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.AAH72793
seq_documentation_block:
ID_AAH72793 standard; cDNA; 1800 BP.
XX
AC_AAH72793;
XX
XX
DT_19-SEP-2001 (first entry)
XX
DE_Human cervical cancer marker nucleic acid 4067.
XX
KM_Cervical cancer: cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS_Homo sapiens.
XX
PN_WO200142467-A2.
XX
PD_14-JUN-2001.
XX
PF_08-DEC-2000; 2000WO-US33312.
XX
PR_08-DEC-1999; 99US-0169681.
PR_21-DEC-1999; 99US-0171350.
PR_14-MAR-2000; 2000US-0189315.
PR_12-MAY-2000; 2000US-0203791.
PR_09-JUN-2000; 2000US-0210600.
PR_21-JUL-2000; 2000US-0220114.
XX
PA_(MIL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI_Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR_WPI; 2001-375006/39.
XX
PT_New isolated nucleic acid for diagnosing and treating cervical cancer
PT_and for assessing and detecting compounds for treating the cancer -
XX
PS_Claim 1: Page 855-856; 1051pp; English.
XX
XX
CC_The invention relates to novel genes (AAH68727-AAH73383) associated with
CC_cervical cancer with cytostatic activity. The nucleic acids and encoded
CC_polypeptides are useful: to assess if a patient is afflicted with
CC_cervical cancer or has a pre-malignant condition; to monitor the
CC_progression of cervical cancer or a premalignant condition in a patient;
CC_to select and/or assess the efficacy of a compound or therapy for
CC_inhibiting cervical cancer in a patient. The nucleic acids may also be
CC_useful for gene therapy.
XX
SO_Sequence 1800 BP; 578 A; 359 C; 466 G; 394 T; 3 other;

alignment_scores:
Quality: 310.00 Length: 553
Ratio: 0.564 Gaps: 1
Percent Similarity: 99.458 Percent Identity: 99.458

alignment_block:
US-09-757-781-2 x AAH72793 ..
Align seg 1/1 to: AAH72793 from: 1 to: 1800

472 GlyPheSerIleThrSerArgAspValThrIleGlyGlySerAlaProI 488
|||||
4 GGATTCACACATCACTTCCAGAGATGTAAACAATGAGTGGCTCAGTCCGAT 53
|||||
488 eTyrValLysAsnIleLeuProArgGlyValAlaIleGlnAspGlyArgL 505
|||||
54 CTATGTGAAAAACATTTCTCCCGGGGGGGGCCATTCAAGATGGCGGAC 103
|||||
505 eulysAlaGlyAspArgLeuIleGluValAsnGlyValAspLeuValGly 521
|||||
104 TTAAGGCAAGAGACAGACTTATAGAGGTAAATGAGTGAATTAATTGGGC 153

```

```

522 LysSerGlnGluValSerLeuArgSerThrLysMetGlu 538
154 AAATCCCAAGAGAGTTGTTCCGTTGACAGACCAAGATGGAAG 203
538 YThrValSerLeuValPheArgGlnGluAspAlaPheHisProArg 555
204 AACTGAGAGCTTGTGCTTTCGACAGAGAGAGCGCTTCCACCCAGAGG 253
555 LuLeuAsnAlaGluProSerGlnMetGlnIleProLysGluThrLysAla 571
254 AACTGAATGACAGACCAAGCCAGATGCGAATTCCAAAGAAAGCAAGCA 303
572 GluAspGluAspIleValLeuThrProAspGlyThrArgGluPheLeu 588
304 GAAGATGAGAGATATTGTTCTTACACCTGATGCGACAGGGAATTTCTGAC 353
588 rPheGluValProLeuAsnAspSerGlySerAlaGlyLeuGlyValSerV 605
354 ATTTGAAGTCCCACTTATGATTCAGATCTGCGAGGCTTGGTGTCAGTG 403
605 allYsGlyAsnArgSerLysGluAsnHisAlaAspLeuGlyIlePheVal 621
404 TCAAAGGATACCGGTCAAAAGAGAACCCAGATTTGGGAATCTTGTGTC 453
622 LysSerIleIleAsnGlyAlaAlaSerLysAspGlyArgLeuArgVa 638
454 AAGTCCATTATTATGAGAGAGAGCATCTAAAGATGGAAGGCTTGCGGT 503
638 LAsnAspGlnLeuIleAlaValAsnGlyLysLeuLeuGlyLysThrA 655
504 GAATGATCAACGTATACGATGAATGGAATCCCTGTGGGACAGACAA 553
655 snGlnAspAlaMetGluThrLeuArgArgSerMetSerThrGlnGlyAsn 671
554 ACCAATATGCCATGAGAAACCTTAAGAGTCTATGTCTACTGAAAGCCAT 603
672 LysArgGlyMetIleGlnLeuIleValAlaArgArgIleSerLysCysAs 688
604 AACGAGAGATGATCCAGCTTATGTTCTCCAAAGAAATTAAGCAAGTGA 653
688 nGluLeuLysSerProGlySerProGlyProGluLeuProIleGlu 705
654 TGAAGTGAAGTCACTGGGAGCCCCCGAGACCTGAGTGGCCATTGAA 703
705 hrAlaLeuAspAspArgGluArgArgIleSerHisSerLeuYrSerGly 721
704 CAGCGTTGATCATAGAGAACGAATTTCCCATTCCTCTACAGTGGG 753
722 IleGluGlyLeuAspGluSerProSerArgAsnAlaIleLeuSerArgI 738
754 ATTCAGGGGCTTGATGAATGCCACAGCAAAATCCCTCAGTAGGAT 803
738 eMetGlyGluSerGlyLysThrGlnLeuSerProThrValAsnMetProG 755
804 AATG.....GTTAAATACAGAGCTGTCCTTACAGTGAATATGCCCC 844
755 LAspAspThrValIleIleGluAspAspArgLeuProValLeuProPro 771
845 AAGATGACACTGATTAAGAGATGACAGGTGCGCATCTCTCCACA 894
772 HisLeuSerAspGlnSerSerSerSerHisAspAspValGlyPheVa 788
895 CACTCTCTACAGTCTCTCTTCCAGCTCCCATGATGATGTTGGGTTTGT 944
788 LThrAlaAspAlaGlyThrTrpAlaIleAlaIleSerAspSerAlaA 805
945 GAGGAGAGATGCTGGTACTTGGGCCAAGGCTGCAATCATGATTCAGCGG 994
805 sPySerSerLeuSerProAspValAspProValLeuAlaPheGlnArgGlu 821
995 ACTGCTCTTGAGTCCAGATGTTGATCCAGTTCGCTTTCAACGAGAA 1044
822 GlyPheGlyArgGlnSerMetSerGluLysArgThrLysGlnPheSerAs 838

```

```

1045 GGAATTGGACGTACAGATATGTCAGAAAAAGCCAAAGCAATTTTCA 1094
838 pAlaSerGlnLeuAspPheValLysThrArgLysSerLysSerMetAspL 855
1095 TTCAGATCAATTTGGATTTCTGTTAAACACGAAATCAAAAAAGCATG 1144
855 euGlyIleAlaAspGluThrLysLeuAsnThrValAspAspGluLysAla 871
1145 TAGGTATAGCTGACAGAGACTAACTCAATACAGTGCATGACGAAAGA 1194
872 GlySerProSerArgAspValGlyProSerLeuGlyLeuLysSerSe 888
1195 GTTCTCTCCACAGAGATGTTGGGTCTCTCCCTGGCTGTAAGATGCAAG 1244
888 rSerLeuGluSerLeuGlnThrAlaValAlaGluValThrLeuAsnGlyA 905
1245 CTCGTTGAGAGTCTCAGACCGCAGTTCGCCAGGTGACTTTGAATGGG 1294
905 sPleProPheHisArgProArgProArgIleIleArgGlyArgGlyCys 921
1295 ATATCTCTTCCATCTCCACGCGCGGATATATCAGAGCAGGGGATGC 1344
922 AsnGluSerPheArgAlaAlaIleAspLysSerYrAspLysProAlaVa 938
1345 AATGAGAGGCTTCAGACCTGCCATCGAACAAATCTTATGATTAACCGG 1394
938 LAspAspAspAspGluGluMetGluThrLeuGluGluLysAspThrGlu 955
1395 AGATGATGATGATGAAGGATGAGACCTTGGAAGAGACACAGAGAA 1444
955 eSerArgSerGlyArgGluSerValSerThrAlaSerAspGlnProSer 971
1445 GTTCAAGATCAGGAGAGAGTGTATTCACAGCCAGTATGATCC 1494
972 HisSerLeuGluArgGlnMetAsnGlyAsnGlnGluLysGlyAspLys 988
1495 CACTCTGAGAGACAAATATGGAACCAAGAGAAAGTATGAAGAC 1544
988 rAspArgGlyLysAspLysThrGlyLysGluLysLysLysAspArgAspL 1005
1545 TGATGAAAAAAGATTAACCTGGAAGAAAGAAAGAAAGATAGAGATA 1594
1005 YsGluLysAspLysMetLysAlaLysGlyMetLeuLysGlyLeuGly 1021
1595 AGGAGAAAGATTAATAATGAAGCCAAAGAGGAATGCTGAAGGCTTGGGA 1644
1022 AspMetPhe 1024
1645 GACATGTTTC 1653

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: AAS27320
seq_documentation_block:
  AAS27320 standard; cDNA: 823 BP.
  AAS27320;
  07-NOV-2001 (first entry)
  DE cDNA encoding novel signal transduction pathway protein, Seq ID 355.
  KW antinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
  KW immune system disorder; rheumatoid arthritis; inflammatory condition;
  KW organ transplant rejection; infection; hepatitis C; blood disorder;
  KW sickle cell anemia; hyperproliferative disorder; Gaucher's disease;
  KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
  KW chromosomal abnormality; Down syndrome; ischemia; renal disorder;
  KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
  KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;
  KW acquired immune deficiency syndrome.

```

OS Homo sapiens.  
XX  
PN WO200154733-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01312.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0186874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225457.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246533.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM  
PI  
XX  
DR WPI: 2001-465460/50.  
P-PSDB; AAU17403.

Claim 1; SEQ ID No 355; 880pp; English

```
alignment_scores:
```

alignment\_block: US-09-757-781-2 x AAS27320 . .

323 CysIleValArgIleAsnAspGlyAspLeuArgAsnArgArgPheGluG1 339

423 HisserAlahisProsestglya  
495 CATAGCGCACACCCCTCGGGAAACACCATCGCTCCAGCCTGGCACC 54

seq\_documentation\_block: 823 BP.

XX  
DT 03-SEP-2001 (first entry)

DE	Human colon cancer	1	Colon cancer antigen: diagnosis; detection
XX			

AA  
OS

Homo sapiens

XX  
PD 05-APR-2001.

XX 29-SEP-1999; 99US-0157137  
 PP

XX  
HUMAN GENOME SCI INC.

XX 00-357 73A

XX  
PT Nucleic acids en

Claim 1; Page 2675; 9803pp; English

CC AAH37195 and AAG7788 represent human colon  
PS cancer-associated nucleic acid molecules (N) and proteins (P), where  
XX AAH37943 to AAH37195 and AAG7788 represent human colon  
CC cancer associated nucleic acid molecules (N) and proteins (P). The colon  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate p  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of p by expressing  
CC inactive proteins or to supplement the patients own production of p.  
CC Additionally, N may be used to produce the colon cancer-associated ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell.  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAH77789 represent sequences used in the exemplification of the

CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX  
SQ Sequence 823 BP; 249 A; 183 C; 201 G; 188 T; 2 other:

## alignment\_scores:

Quality: 195.00 Length: 195  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-757-781-2 x AAF33594 ..

Align seg 1/1 to: AAF33594 from: 1 to: 823

```
323 CysIleValArgIleAsnAspGlyAspLeuArgAsnArgPheGluG1 339
|||||
195 TGCATGTCAGAGTAAATGATGCGCGACCTTCGAATAGAAATGTTGAAAC 244
339 nAlaGlnHisMetPheArgGlnAlaMetArgThrProIleIleThrPheH 356
|||||
245 AGCAACAATATGTTGTCGCAAGCCATGCGTACACCCATCATTTGGTTC 294
356 IsValValProAlaAlaAsnLysGluGlnIleGlnIleSerGlnSer 372
|||||
295 ATGGGTTCCTCGACGAAATAAAGACAGTATGAACAACATATCCCAAGT 344
373 GluLysAsnAsnIleTyrSerSerArgPheSerProAspSerGlnTyrI1 389
|||||
345 GAGAAAGAACATTAATCTATTCAGCCGTTTAAAGCCCTGACACCCAGTATAT 394
389 eAspAsnArgSerValAsnSerAlaGlyLeuHisThrValGlnArgAlaP 406
|||||
395 TGACAAACAGAGGTGTGAACAGTGCAGGCTTCAACAGGTGACAGAGCAC 444
406 roATgLeuAsnHisProProGluGlnIleAspSerHisSerArgLeuPro 422
|||||
445 CCCAGTGAACACACCCGCTGAGCAGATAGACTCTCACTCAAGACTACCT 494
423 HisSerAlaHisProSerGlyLysProProSerAlaProAlaSerAlaP 439
|||||
495 CATAGCCCAACACCCCTCGGAAACACACATCCGCTCAGCCGCGCAC 544
439 oGlnAsnValPheSerThrThrValSerSerGlyTyrAsnThrIleLysI 456
|||||
545 TCAGAAATGATTTAGTACGACTGTAACAGTGGTTATACCCAAAAA 594
456 leGlyLysArgLeuAsnIleGlnLeuLysGlyThrGluGlyLeuGly 472
|||||
595 TAGCAAGAGGCTTAATATCCAGCTTAAGAAAGTACGAAGGTTGGGA 644
473 PheSerIleThrSerArgAspValThrIleGlyLysSerAlaProIleTyr 489
|||||
645 TTCAGCATCACTTCACAGATGTAACAATAGTGCTCAGCTCCAACTTA 694
489 rValLysAsnIleLeuProArgGlyAlaAlaIleGlnAspGlyArgLeuL 506
|||||
695 TGTGAAAAACATTTCTCCCGCGGGGCGGCGCATTCAGGATGGCGACTTA 744
506 ysaLagIysArgLeuIleGluValAsnGlyVal 517
|||||
745 AGCGAGAGACAGACTATAGAGTAAATGAGTA 779
```

seq\_name: /STDsJ/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF67583

seq\_documentation\_block:

ID AAF67583 standard; cDNA; 425 BP.

XX  
AC AAF67583;  
XX

DT 09-APR-2001 (first entry)  
XX  
DE Novel human polynucleotide, SEQ ID NO: 3345.  
XX  
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;  
KW breast cancer; lung cancer; cancer detection; ss.  
XX

OS Homo sapiens.

PN WO200102568-A2.

PD 11-JAN-2001.

XX 30-JUN-2000; 2000MO-US18374.

XX 02-JUL-1999; 99US-0142310.

XX 02-JUL-1999; 99US-0142311.

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;  
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;  
PI Ckenjakov R, Drmanac S, Dickson M, Labat I, Leschkowitz D;  
PI Kita D, Garcia V, Jones LW, Strache-Crain B;

XX WPI; 2001-091805/10.

DR

XX

PT

PT

XX

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

## alignment\_block:

US-09-757-781-2 x AAF67583 ..

Align seg 1/1 to: AAF67583 from: 1 to: 425

## alignment\_scores:

Quality: 131.00 Length: 131  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

```
586 PheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGlyLeuG1 602
|||||
33 TTTCGACATTTGAAGTCCCACTTAATATGATTCAGATCTGCAAGCTTGG 82
602 yValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLeuGlyI 619
|||||
83 TGTGAGTGTCAAGGTACCGGTCAAAAGAGAACACCAAGATTTGGGAA 132
619 lePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspGlyArg 635
|||||
```

```

alignment_scores:
  Quality: 110.00      Length: 110
  Ratio: 1.000
  Percent Similarity: 100.000
  Gaps: 0
  Percent Identity: 100.000
alignment block:
msc00_75

```

From: 1 to: 2554

```
seq_name: /S1D51/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:AA568903
seq_documentation_block:
ID      AA568903 standard; cDNA; 1262 BP.
XX
```

1002 (first entry)  
DNA encoding novel human diagnostic protein #4707,  
Human; chromosome 1

PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	

PF 30-MAR-2001; 2001WO-US08631.





Thu Jul 25 08:38:21 2002

us-09-757-781-2.oli.g.rng

```
468 ThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly1 484
|||||
1 ACAGAGGTTGGGATTCAGCATCTCCAGATGTAATAGAGTGG 50
|||||
484 ySerAlaProIleTyrValIysAsnIleLeuProArgIValAlaIleG 501
|||||
51 CTGAGCTCCAAATCTATGTGAAAACATTCCTCCCGGGGGCCGCAATTC 100
|||||
501 InaSpGlyArgLeuIysAlaGlyAspArgLeuIleGlyValAsnGly1 517
|||||
101 AGGATGGCCCGACTTAAGGCGAGAGACTTATAGAGTAATGAGACTA 150
|||||
518 AspleuValGlyIysSerGlnGluValValSerLeuLeuArgSerTh 534
|||||
151 GATTAGTGGGCAAAATCCCAAGAGAGAGTGTTCCTTCCCTTGAGAAGC 200
|||||
534 rlyMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 551
|||||
201 CAAGATGGAGAGACTGTGAGCTTCTGTCTTCCGCGAGAGAGCCCT 250
|||||
551 heHisProArgGluLeu 556
|||||
251 TCCACCCAGAGGAACTG 267
|||||

seq_name: /SIDSL/gcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF66057
seq_documentation_block:
ID AAF66057 standard; cDNA: 400 BP.
XX
AC AAF66057;
XX
DF 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 1813.
XX
DE Human: cytostatic; gene therapy; colon cancer; prostate cancer;
XX
KW breast cancer; lung cancer; cancer detection; ss.
XX
OS Homo sapiens.
XX
PN WO200102568-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US18374.
XX
PR 02-JUL-1999; 99US-0142310.
XX
PR 02-JUL-1999; 99US-0142311.
XX
PA (CHIR ) CHIRON CORP.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Cirenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones IM, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
XX
PT mammalian cell and detecting cancer, particularly of the colon or
XX
PT prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9; Page 803; 1046pp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
XX
CC polynucleotides. The library is used to detect differentially expressed
XX
CC genes correlated with a cancerous state of a mammalian cell and can
XX
CC detect colon, prostate, breast and lung cancer. The library can be used
XX
CC to produce probes for detection of mRNA and to produce additional copies
XX
CC of the polynucleotides. The probes can be used for chromosome mapping of
XX
CC the polynucleotide and for detection of transcription levels. Ribozymes
XX
CC or antisense oligonucleotides can be generated. The polynucleotides and
```

CC their gene products are used as genetic or biochemical markers (e.g. in  
CC blood or tissues) that will detect the earliest changes along the  
CC carcinogenesis pathway and/or monitor the efficacy of therapies and  
CC preventive interventions. The polynucleotides, polypeptides and  
CC antibodies against them can be used in pharmaceutical compositions to  
CC treat the cancers and proliferative disorders such as neoplasia,  
CC dysplasia and hyperplasia.

SO Sequence 400 BP; 118 A; 85 C; 99 G; 98 T; 0 other;

alignment\_scores:                    Length:       84  
                  Quality:               0  
Percent Similarity: 100.000       Percent Identity: 100.000

alignment\_block:  
US-09-757-781-2 x AAF66057 ..  
Align seg 1/1 to: AAF66057 from: 1 to: 400

```
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeu1 817
|||||
149 AGTGAATTCAGCCGACTGCTTGTGATCCAGATGTGATCAATCTTCTG 198
|||||
817 apheGlnArgGluGlyPheGlyArgGlnSerMetSerGluIysArgThrL 834
|||||
199 TTTTCAACGAGAGGATTTGAGCTGACAGATATGCGAAAAACGCCAAC 248
|||||
834 ysgInPheSerAspAlaSerGlnLeuAspPheValIysThrArgIysSer 850
|||||
249 AGCAATTTTCAGATGCCAGCAATTTGATTTGTTAAACAGAAATCA 298
|||||
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
|||||
299 AAAAGCATGATTTAGGTATAGCTGACGAGACTAAACTCAATACAGTGA 348
|||||
867 pAspGlnIysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
|||||
349 TGACCAGAAAGCAGGTCTCCACAGACAGATGTGGGTCTTCCTGGGTC 398
|||||
884 eu 884
|||
399 TG 400
```

seq\_name: /SIDSL/gcdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AA568901  
seq\_documentation\_block:  
ID AA568901 standard; cDNA: 582 BP.  
XX  
AC AA568901;  
XX  
DF 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #4705.  
XX  
DE Human: chromosome mapping; gene mapping; gene therapy; forensic;  
XX  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;



seq\_documentation\_block:

ID AAD23370 standard; cDNA; 540 BP.

AC AAD23370;

26-FEB-2002 (first entry)

Human lung tumour-specific SCC1-568 cDNA.

Human: lung tumour protein; immunostimulant; cytostatic; gene therapy; antisense-therapy; vaccine; immune response; lung cancer; SCC1-568; ss.

WO200172295-A2.

04-OCT-2001.

28-MAR-2001; 2001WO-US09991.

29-MAR-2000; 2000US-0538037.

18-AUG-2000; 2000US-0588937.

22-SEP-2000; 2000US-0640878.

01-NOV-2000; 2000US-234517P.

14-DEC-2000; 2000US-0704512.

(CORI-) CORIXA CORP.

Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indrias CY, Henderson RA, Filing SP, Algate PA, Elliot M, Mannion J, Kalos MD;

WPI; 2001-639201/73.

Example 2; Page 276; 378pp; English.

The invention relates to isolated lung tumour-specific proteins and their corresponding cDNA molecules. Lung tumour-specific proteins and their antigen-presenting cells are useful for stimulating and/or the development of cancer. The invention also relates to a composition useful for stimulating an immune response, and for inhibiting lung tumour specific oligonucleotide is useful in gene therapy and for diagnosis, detection and treatment of lung cancer. The present sequence is human lung tumour-specific cDNA.

Sequence 540 BP; 147 A; 123 C; 133 G; 136 T; 1 other;

alignment\_scores:

Quality: 78.00 Length: 78  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment block:

US-09-757-781-2 x AAD23370

Align seg 1/1 to: AAD23370 from: 1 to: 540

743 Glycerol leucine serine proline valine methionine aspartic acid  
229 GGTAAATACAGCTGCTCCCTACAGTGAATATGCCCCAAGATGACACTGT 278  
759 TTTTGGTAAATACAGCTGCTCCCTACAGTGAATATGCCCCAAGATGACACTGT 278  
779 CATTATAGAAAGATGACAGCTGCTCCCTCCACATCTCTGACC 328  
776 TTTTGGTAAATACAGCTGCTCCCTCCACATCTCTGACC 328  
776 TTTTGGTAAATACAGCTGCTCCCTCCACATCTCTGACC 328

329 AGTCTCTTCACAGCTCCATCATGATGTGGGTTTGAGAGGACAGTCT 378  
793 Glycerol leucine serine proline valine methionine aspartic acid  
379 GGTAAATACAGCTGCTCCCTACAGTGAATATGCCCCAAGATGACACTGT 428  
809 TTTTGGTAAATACAGCTGCTCCCTCCACATCTCTGACC 428  
429 TTTTGGTAAATACAGCTGCTCCCTCCACATCTCTGACC 428  
seq\_name: /SIDSL/gcdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH07109

seq\_documentation\_block:

ID AAH07109 standard; cDNA; 556 BP.

AAH07109;

26-JUN-2001 (first entry)

Human cDNA clone (5'-primer) SEQ ID NO:3944.

Human: primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118176.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Claim 1; SEQ ID 3944; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH95893 represent human amino acid sequences; AAH92446 to AAH13628 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Thu Jul 25 08:38:21 2002

XX Sequence 556 BP; 154 A; 129 C; 144 G; 126 T; 3 other;

alignment\_scores:

Quality:	64.00	Length:	64
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment\_block:

US-09-757-781-2 x AAH07109 ..

Align seg 1/1 to: AAH07109 from: 1 to: 556

```
126 GluValThrProSerValIleuArgAlaAsnMetProLeuHisValArgAr 142
|||||
3 GAGGTACACACTTCAGTCCTCGAGCAAAATGGCTCTTCATGTTGAGCG 52
|||||
142 gSerSerAspProAlaLeuIleGlyLeuSerThrServaIserAspSera 159
|||||
53 CAGTAGTGACCCAGCTCTAATTGGCTCTCCACTTGTGTGAGTATAGTA 102
|||||
159 snpSerSerGluGluProSerArgLysAsnProThrArgTrpSerThr 175
|||||
103 ATTTTCTCTGAAAGAGCTTCAAGAAAAATCCACACGCTGGTCAACA 152
|||||
176 ThrAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLys 189
|||||
153 ACAGCTGGCTTCTCAAGCAGACACACTGTGGAGTCTCTAAA 194
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:33:34 ; Search time 30.49 Seconds

(without alignments)  
7693.707 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 1356

Sequence: 1 MKYVCFGRTRVVPCGDH.....SQVRLNRLQTPKGRPEYS 1356

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size: 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1113	82.1	1353	4	Q9BY58 homo sapien
2	781	57.6	1031	4	Q9BY57 homo sapien
3	556	41.0	988	4	Q96RM7 homo sapien
4	556	41.0	1273	4	Q96RM6 homo sapien
5	500	36.9	1266	4	Q9NYE6 homo sapien
6	413	30.5	667	4	Q9HC48 homo sapien
7	406	29.9	865	4	Q96K28 homo sapien
8	347	25.6	347	4	Q9NM14 homo sapien
9	88	6.5	1337	11	Q9Z340 ratu mus
10	77	5.7	1333	11	Q9Z340 ratu mus
11	19	1.4	624	4	Q96DK9 homo sapien
12	19	1.4	1143	4	Q96NM6 homo sapien
13	19	1.4	1205	4	Q96NM7 homo sapien
14	13	1.0	545	4	Q96N09 homo sapien
15	12	0.9	128	6	Q95L72 macaca fasc
16	11	0.8	1464	5	Q96782 drosophila

17	11	0.8	1464	5	Q9YX75	Q9YX75 drosophila
18	11	0.8	2460	11	Q64512	Q64512 mus musculu
19	10	0.7	79	4	Q96Q29	Q96Q29 homo sapien
20	10	0.7	213	11	Q9CSB4	Q9CSB4 mus musculu
21	10	0.7	1017	16	Q9ZMD1	Q9ZMD1 rhizobium m
22	8	0.6	74	6	Q9YX25	Q9YX25 canis famill
23	8	0.6	88	2	Q9ZGY1	Q9ZGY1 yersinia pe
24	8	0.6	107	3	Q9Y813	Q9Y813 schizosacch
25	8	0.6	128	10	Q93020	Q93020 treponema p
26	8	0.6	153	16	Q96120	Q96120 treponema p
27	8	0.6	219	16	Q9K618	Q9K618 bacillus ha
28	8	0.6	226	11	Q9DC86	Q9DC86 mus musculu
29	8	0.6	287	10	Q9ZK23	Q9ZK23 triticum sp
30	8	0.6	296	4	Q14909	Q14909 homo sapien
31	8	0.6	297	5	P90976	P90976 caenorhabd
32	8	0.6	315	11	Q91Y70	Q91Y70 rattus norv
33	8	0.6	316	5	Q9U245	Q9U245 caenorhabd
34	8	0.6	318	5	Q22229	Q22229 caenorhabd
35	8	0.6	323	2	Q930E0	Q930E0 rhizobium m
36	8	0.6	326	4	Q00272	Q00272 homo sapien
37	8	0.6	334	11	Q9D3C6	Q9D3C6 mus musculu
38	8	0.6	335	2	Q9E2T7	Q9E2T7 streptomyce
39	8	0.6	335	11	Q91Y06	Q91Y06 mus musculu
40	8	0.6	336	16	Q92SA7	Q92SA7 rhizobium m
41	8	0.6	337	4	Q00556	Q00556 homo sapien
42	8	0.6	337	11	Q9JHL1	Q9JHL1 mus musculu
43	8	0.6	337	11	Q920G2	Q920G2 rattus norv
44	8	0.6	373	11	Q9CVL4	Q9CVL4 mus musculu
45	8	0.6	450	4	Q15599	Q15599 homo sapien

## ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	1353 AA.
Q9BY58	Q9BY58			
AC	Q9BY58			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN LONG VARIANT.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID-9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Fang C., Xu Y.,			
RT	"Exon/Intron Structure and Splicing Variants of a Novel Human Polarity			
RT	gene, hASIP."			
RL	Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF196185; AAK27891.1; -.			
DR	HSSP: Q12923; 3PDZ.			
DR	InterPro: IPR001478; PDZ.			
DR	Pfam: PF00595; PDZ. 3.			
DR	SMART: SM00228; PDZ. 3.			
DR	PROSITE: PSS0106; PDZ. 3.			
SQ	SEQUENCE 1353 AA; 151149 MW; BC531577B9C31AE3 CRC64;			

Query Match 82.1%; Score 1113; DB 4; Length 1353;  
Best local similarity 99.8%; Pred. No. 0;  
Matches 1353; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY	1	MKYVCFGRTRVVPCGDHMKVFSLIQAVTRYRAIAKDPYKIQVRLHFGDGIID 60
DB	1	MKYVCFGRTRVVPCGDHMKVFSLIQAVTRYRAIAKDPYKIQVRLHFGDGIID 60
QY	61	LDIILCDVADDDKRLVAVDEDDPHHGGDGTSSSTGTOSPEIFGSELGNNVASAPQY 120
DB	61	LDIILCDVADDDKRLVAVDEDDPHHGGDGTSSSTGTOSPEIFGSELGNNVASAPQY 120

```
QY 121 ATSEIVTSPVLNANMPLHVRSSDPALIGLSTSVSSNESSEPSRKNPTRMSTTAGFL 180
Db 121 ATSEIVTSPVLNANMPLHVRSSDPALIGLSTSVSSNESSEPSRKNPTRMSTTAGFL 180
QY 181 KONTAGSPKTCORRKKDEMYRSLPRDTSNWSNOFORDNARSSLASHPMVGKMLEKODE 240
Db 181 KONTAGSPKTCORRKKDEMYRSLPRDTSNWSNOFORDNARSSLASHPMVGKMLEKODE 240
QY 241 DGTEDNSRVEPVGHADTGLEHINPNSLDDMYKLYEVPNNGGGLGHVVPFSAARGRTLG 300
Db 241 DGTEDNSRVEPVGHADTGLEHINPNSLDDMYKLYEVPNNGGGLGHVVPFSAARGRTLG 300
QY 301 LLYKLEKGGKAHEHENLFRENDCTIVRINDGDLRNRREFEOAHMFROAMRTPIIMFHVPA 360
Db 301 LLYKLEKGGKAHEHENLFRENDCTIVRINDGDLRNRREFEOAHMFROAMRTPIIMFHVPA 360
QY 361 ANKEQYEQLSQSEKNKNTYSSRFSPOQYIDNRSVNSAGLHYVQAPRLNHPPEQIDSHR 420
Db 361 ANKEQYEQLSQSEKNKNTYSSRFSPOQYIDNRSVNSAGLHYVQAPRLNHPPEQIDSHR 420
QY 421 LPHSAHPSGKPPASAPASAPQNVFTTVSSGYNTKKIGKRLNQLKKTGEGIGFTSRDY 480
Db 421 LPHSAHPSGKPPASAPASAPQNVFTTVSSGYNTKKIGKRLNQLKKTGEGIGFTSRDY 480
QY 481 TIGGSAPITYKNTIPRGAALIODGRKAGBRLIEVNGVDLYGKSOEYVSLSTRKMGTV 540
Db 481 TIGGSAPITYKNTIPRGAALIODGRKAGBRLIEVNGVDLYGKSOEYVSLSTRKMGTV 540
QY 541 SILVFEQEDAFHPRELNEPSPOMQIPKTKAEDIEDIVLTPDGTREFLTFEVPINDSSAG 600
Db 541 SILVFEQEDAFHPRELNEPSPOMQIPKTKAEDIEDIVLTPDGTREFLTFEVPINDSSAG 600
QY 601 LGVSVKGNRSKENHADLGIKFKSIINGGAASKOGRRLRVNOLIAVNGESLGTNODAME 660
Db 601 LGVSVKGNRSKENHADLGIKFKSIINGGAASKOGRRLRVNOLIAVNGESLGTNODAME 660
QY 661 TLRRSMSTEGKRGKIOIIVARRISKNEKSPGSPPELPETALDRERISHSLYS 720
Db 661 TLRRSMSTEGKRGKIOIIVARRISKNEKSPGSPPELPETALDRERISHSLYS 720
QY 721 GIBELDSSPSNNAALSTRIMESGKYQLSPVNNPQDDYIIEEDRLPYLPHLSQSSSS 780
Db 721 GIBELDSSPSNNAALSTRIMESGKYQLSPVNNPQDDYIIEEDRLPYLPHLSQSSSS 780
QY 781 SHDVGFTVADAGTMAKAAISDSADCSLSPDVPVLAFOREGFGQSMSEKRTQFSDAS 840
Db 781 SHDVGFTVADAGTMAKAAISDSADCSLSPDVPVLAFOREGFGQSMSEKRTQFSDAS 840
QY 841 QIDPVYTRKSKMDLGIADETKLTNTVDQKAGSPRDYGPGLGKSSLSLQTAFAEV 900
Db 841 QIDPVYTRKSKMDLGIADETKLTNTVDQKAGSPRDYGPGLGKSSLSLQTAFAEV 900
QY 901 TLNGDIPFRRPRRIIRGCGNESFRAALIDSYDKPAVDDDEGEMTLEEDTRESSSGR 960
Db 901 TLNGDIPFRRPRRIIRGCGNESFRAALIDSYDKPAVDDDEGEMTLEEDTRESSSGR 960
QY 961 EASYSTRASDOPSHLERONGNOEGKDTDRKKDKTGKKEKKDKRDEKDKAKKMGKGL 1020
Db 961 EASYSTRASDOPSHLERONGNOEGKDTDRKKDKTGKKEKKDKRDEKDKAKKMGKGL 1020
QY 1021 GDMFEFGKRRKDKTEKGIKTIQESTSEEBERIRMKOEQERIOAKTREFREKQARERDY 1080
Db 1021 GDMFEFGKRRKDKTEKGIKTIQESTSEEBERIRMKOEQERIOAKTREFREKQARERDY 1080
QY 1081 AEIODFHTFECDDLELMYGVSVSESGMALNARPOSPREGHMDALYAAVKKPRNSKSP 1140
Db 1081 AEIODFHTFECDDLELMYGVSVSESGMALNARPOSPREGHMDALYAAVKKPRNSKSP 1140
QY 1141 VDSNRSTPSNNDRIQRLROFEOAKODEYEDRRRTYSEFOQWPNARPAITOGGRISVAYE 1200
Db 1141 VDSNRSTPSNNDRIQRLROFEOAKODEYEDRRRTYSEFOQWPNARPAITOGGRISVAYE 1200
QY 1201 VQMKROQOEERESSQAQROYSLPROSKNASSVSQDSWEDNTYSPGEGFOSAKENPRTS 1260
Db 1201 VQMKROQOEERESSQAQROYSLPROSKNASSVSQDSWEDNTYSPGEGFOSAKENPRTS 1260
```

```
Db 1198 VQMKROQOEERESSQAQROYSLPROSKNASSVSQDSWEDNTYSPGEGFOSAKENPRTS 1257
QY 1261 SYGSRNGYIGGHNARVNLLETQELLROBORRKEQOMKOPSEGPSNYSYKKVODPS 1320
Db 1258 SYGSRNGYIGGHNARVNLLETQELLROBORRKEQOMKOPSEGPSNYSYKKVODPS 1317
QY 1321 YAPPKGPFRODVPSPSQVARNLRLQTPPEKGRPFS 1356
Db 1318 YAPPKGPFRODVPSPSQVARNLRLQTPPEKGRPFS 1353
```

```
RESULT 2
Q9BY57 PRELIMINARY; PRT: 1031 AA.
ID 09BY57;
AC 09BY57;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
OS Homo sapiens (Human).
OC Homo sapiens (Human).
OC Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Fang C., Xu Y.;
RT "Exon/Intron Structure and Splicing Variants of a Novel Human Polarity
RT Gene, hASIP.";
RL Submitted (0CT-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AP196186; AAK27892.1; -.
DR HSRP: Q12923; 3PZD;
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 3.
DR SMART: SM00228; PDZ; 3.
DR PROSITE: PS50106; PDZ; 3.
SQ
SEQUENCE 1031 AA; 113418 MW; 40DD593636EAE999 CRC64;
```

Query Match Best Local Similarity 57.6%; Score 781; DB 4; Length 1031; Matches 1021; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

```
QY 1 MMYTCFGFTRVYVPCGGGHHKVFSLIOQAVTRYRKAIANDPNYWIQVHLEHGGDGIID 60
Db 1 MMYTCFGFTRVYVPCGGGHHKVFSLIOQAVTRYRKAIANDPNYWIQVHLEHGGDGIID 60
QY 61 LDDILCDVAADKRLVAVFDEQDPHHGGDGTSASTGTOSPELFGSELGTNNVSAFOPYO 120
Db 61 LDDILCDVAADKRLVAVFDEQDPHHGGDGTSASTGTOSPELFGSELGTNNVSAFOPYO 120
QY 121 ATSEIVTSPVLNANMPLHVRSSDPALIGLSTSVSSNESSEPSRKNPTRMSTTAGFL 180
Db 121 ATSEIVTSPVLNANMPLHVRSSDPALIGLSTSVSSNESSEPSRKNPTRMSTTAGFL 180
QY 181 KONTAGSPKTCORRKKDEMYRSLPRDTSNWSNOFORDNARSSLASHPMVGKMLEKODE 240
Db 181 KONTAGSPKTCORRKKDEMYRSLPRDTSNWSNOFORDNARSSLASHPMVGKMLEKODE 240
QY 241 DGTEDNSRVEPVGHADTGLEHINPNSLDDMYKLYEVPNNGGGLGHVVPFSAARGRTLG 300
Db 241 DGTEDNSRVEPVGHADTGLEHINPNSLDDMYKLYEVPNNGGGLGHVVPFSAARGRTLG 300
QY 301 LLYKLEKGGKAHEHENLFRENDCTIVRINDGDLRNRREFEOAHMFROAMRTPIIMFHVPA 360
Db 301 LLYKLEKGGKAHEHENLFRENDCTIVRINDGDLRNRREFEOAHMFROAMRTPIIMFHVPA 360
QY 361 ANKEQYEQLSQSEKNKNTYSSRFSPOQYIDNRSVNSAGLHYVQAPRLNHPPEQIDSHR 420
Db 361 ANKEQYEQLSQSEKNKNTYSSRFSPOQYIDNRSVNSAGLHYVQAPRLNHPPEQIDSHR 420
QY 421 LPHSAHPSGKPPASAPASAPQNVFTTVSSGYNTKKIGKRLNQLKKTGEGIGFTSRDY 480
Db 421 LPHSAHPSGKPPASAPASAPQNVFTTVSSGYNTKKIGKRLNQLKKTGEGIGFTSRDY 480
```



```
Db 421 LPHSAHPGKPPSAPASAPQNVSTTVSSGYNTKTKIGKRLNIOCLKGTGCGFSITSRDV 480
Qy 481 TIGGSAPIYVKNIPLPGAAIIOGRLKAGDRLIEVNGVDLVGKSOEEVSLIIRSTKMGTV 540
Db 481 TIGGSAPIYVKNIPLPGAAIIOGRLKAGDRLIEVNGVDLVGKSOEEVSLIIRSTKMGTV 540
Qy 541 SILVFRQEDAFHPRRLNAEPQOMQIPKETYKADEDEDIVLPDGTREFLTFEVPPLNDGSAG 600
Db 541 SILVFRQEDAFHPRRLNAEPQOMQIPKETYKADEDEDIVLPDGTREFLTFEVPPLNDGSAG 600
Qy 601 LCVSVKGNSSKKNHMDLGIFFVKSIIINGGAASKDGRLVNDQILAVNGESLIGTNDADAE 660
Db 601 LCVSVKGNSSKKNHMDLGIFFVKSIIINGGAASKDGRLVNDQILAVNGESLIGTNDADAE 660
Qy 661 TLRRSMTSGNRGMQILIVARRISKCNELKSGSPGPELPIETALDRERRISHSIYS 720
Db 661 TLRRSMTSGNRGMQILIVARRISKCNELKSGSPGPELPIETALDRERRISHSIYS 720
Qy 721 GIEGLDESPSRNAALSRIINGESGKYQLSPTVNMPODDTVIIEDDRPLVLPPLHSDGSSSS 780
Db 721 GIEGLDESPSRNAALSRIIM---GKYQLSPTVNMPODDTVIIEDDRPLVLPPLHSDGSSSS 777
Qy 781 SHDDVGFVYADAGTWAKAIIISDACSLSPDVDPVLAFOREGGRGSMSEKRTKQFSDDS 840
Db 778 SHDDVGFVYADAGTWAKAIIISDACSLSPDVDPVLAFOREGGRGSMSEKRTKQFSDDS 837
Qy 841 QLDFFVTKRSKSMDLGIADEFTKLNVDQKAGSPSRDVPGLKSSLSLSTQTAFAVY 900
Db 838 QLDFFVTKRSKSMDLGIADEFTKLNVDQKAGSPSRDVPGLKSSLSLSTQTAFAVY 897
Qy 901 TLNGDIPFHRPRLIIRGRGNESEFRAAIDKSYDKPAVDDDDGEMTLEEDTESSRSGR 960
Db 898 TLNGDIPFHRPRLIIRGRGNESEFRAAIDKSYDKPAVDDDDGEMTLEEDTESSRSGR 957
Qy 961 EESVSTASDQPSLSLEOMNGNCKGDKTKKDKTKGKKKKDKKKDKKAKKMLKGL 1020
Db 958 EESVSTASDQPSLSLEOMNGNCKGDKTKKDKTKGKKKKDKKKDKKAKKMLKGL 1017
Qy 1021 GDMF 1024
Db 1018 GDMF 1021

RESULT 3
Q96RM7 PRELIMINARY: PRT: 988 AA.
AC 096RM7:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN SHORT VARIANT
DE B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pang C., Xu Y.;
RT "Exon/Intron Structure And Splicing Variants of A Novel Human Polarity
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF332593; AAK69192.1; -.
SQ SEQUENCE 988 AA; 108545 MW; 89F2139B096F7F7E CFC64;
```

Query Match 41.0%; Score 556; DB 4; Length 988;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MKYTVCFGRTRVVPCGDDGIMKVFSLIOQAVTRRKAIADPNYWIQVHRLHGGDGLD 60
Db 1 MKYTVCFGRTRVVPCGDDGIMKVFSLIOQAVTRRKAIADPNYWIQVHRLHGGDGLD 60
```

```
Qy 61 LDDILCDVADDKRLVAVDEQDPHHGGDGTSSSTGTOSPEIFGSELGTNNVSAPQPY 120
Db 61 LDDILCDVADDKRLVAVDEQDPHHGGDGTSSSTGTOSPEIFGSELGTNNVSAPQPY 120
Qy 121 ATSEIEVTSVLANPPLHVRSSDPALITGLSTSVSDSNFSSEEPSKKNPTKSTTAGFL 180
Db 121 ATSEIEVTSVLANPPLHVRSSDPALITGLSTSVSDSNFSSEEPSKKNPTKSTTAGFL 180
Qy 181 KONTAGSPKTCDDKKNENRSLPRDTSNMSNOQORANRSSLASHPMVKMLEKODE 240
Db 181 KONTAGSPKTCDDKKNENRSLPRDTSNMSNOQORANRSSLASHPMVKMLEKODE 240
Qy 241 DGTEDNSRVEPYGHADTGLEHINPFLSDMVAFLVEPNDCGLGIHVVPFSARGRTLG 300
Db 241 DGTEDNSRVEPYGHADTGLEHINPFLSDMVAFLVEPNDCGLGIHVVPFSARGRTLG 300
Qy 301 LLYKRLKGGKAEHENLFRENDCIVRINDGDLNRRREQAQMFRQAMRTPIIFHVPA 360
Db 301 LLYKRLKGGKAEHENLFRENDCIVRINDGDLNRRREQAQMFRQAMRTPIIFHVPA 360
Qy 361 ANKEQYEOLOSSEKNNYSSRSPDSQYIDNRVNSAGLHTVQRAPLNNHPPQIOSHR 420
Db 361 ANKEQYEOLOSSEKNNYSSRSPDSQYIDNRVNSAGLHTVQRAPLNNHPPQIOSHR 420
Qy 421 LPHSAHPGKPPSAPASAPQNVSTTVSSGYNTKTKIGKRLNIOCLKGTGCGFSITSRDV 480
Db 421 LPHSAHPGKPPSAPASAPQNVSTTVSSGYNTKTKIGKRLNIOCLKGTGCGFSITSRDV 480
Qy 481 TIGGSAPIYVKNIPLPGAAIIOGRLKAGDRLIEVNGVDLVGKSOEEVSLIIRSTKMGTV 540
Db 481 TIGGSAPIYVKNIPLPGAAIIOGRLKAGDRLIEVNGVDLVGKSOEEVSLIIRSTKMGTV 540
Qy 541 SILVFRQEDAFHPRREL 556
Db 541 SILVFRQEDAFHPRREL 556
```

```
RESULT 4
Q96RM6 PRELIMINARY: PRT: 1273 AA.
AC 096RM6:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN LONG VARIANT
DE B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pang C., Xu Y.;
RT "Exon/Intron Structure And Splicing Variants of A Novel Human Polarity
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF332593; AAK69193.1; -.
SQ SEQUENCE 1273 AA; 141730 MW; 53C1A94D8CB7341E CFC64;
```

Query Match 41.0%; Score 556; DB 4; Length 1273;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MKYTVCFGRTRVVPCGDDGIMKVFSLIOQAVTRRKAIADPNYWIQVHRLHGGDGLD 60
Db 1 MKYTVCFGRTRVVPCGDDGIMKVFSLIOQAVTRRKAIADPNYWIQVHRLHGGDGLD 60
Qy 61 LDDILCDVADDKRLVAVDEQDPHHGGDGTSSSTGTOSPEIFGSELGTNNVSAPQPY 120
Db 61 LDDILCDVADDKRLVAVDEQDPHHGGDGTSSSTGTOSPEIFGSELGTNNVSAPQPY 120
```

```
QY 121 ANSEIEVTPSVLRANMPLHVRNRSSDPALIGLSTVSDSNFSESPSRKKNPTRMSTTAGFL 180
DB 121 ANSEIEVTPSVLRANMPLHVRNRSSDPALIGLSTVSDSNFSESPSRKKNPTRMSTTAGFL 180
QY 181 KONTAGSPTCKDKKDEKENTRSLPRTSNMNSNOFORNASSSLASHPMYGKMLEKODE 240
DB 181 KONTAGSPTCKDKKDEKENTRSLPRTSNMNSNOFORNASSSLASHPMYGKMLEKODE 240
QY 241 DGTEDNSREVEPYGHADTGLEHINPNSLDDMYLVEVPNDGPGLGTHVVPESARGRTIG 300
DB 241 DGTEDNSREVEPYGHADTGLEHINPNSLDDMYLVEVPNDGPGLGTHVVPESARGRTIG 300
QY 301 LUYKLEKCKGKAHEMLFPENDCIYRINDGLRNRRFEQAQHMFRQAMPTPIIWEHVPA 360
DB 301 LUYKLEKCKGKAHEMLFPENDCIYRINDGLRNRRFEQAQHMFRQAMPTPIIWEHVPA 360
QY 361 ANKEQEOJLSEKKNYSSRFSPODYIDNRSVNSAGLHTVORARLHHPPEQIDSHSR 420
DB 361 ANKEQEOJLSEKKNYSSRFSPODYIDNRSVNSAGLHTVORARLHHPPEQIDSHSR 420
QY 421 LPHSAHPSGKPPSAPASAPQNFSTVSSGYNTKKIGKRLNIQKKTEGLGFSITSRDY 480
DB 421 LPHSAHPSGKPPSAPASAPQNFSTVSSGYNTKKIGKRLNIQKKTEGLGFSITSRDY 480
QY 481 TTGGSPITYVKNIIDRGAAIIDGRLKACDRLIEVNGVDLVGKSQOEVSLSLSTMEGTIV 540
DB 481 TTGGSPITYVKNIIDRGAAIIDGRLKACDRLIEVNGVDLVGKSQOEVSLSLSTMEGTIV 540
QY 541 SILVFRQEDAFHPRRL 556
DB 541 SILVFRQEDAFHPRRL 556

RESULT 5
Q9NVE6 PRELIMINARY: PRT: 1266 AA.
ID 09NVE6:
AC 09NVE6:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PAR3.
GN PAR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE=20394296; PubMed=10934474;
RA Joberty G., Petersen C., Gao L., Macara I.G.;
RT "The cell-polarity protein Par6 links Par3 and atypical protein kinase
RT C to Cdc42."
RL Nat. Cell Biol. 2:531-539(2000).
DR EMBL: AF252293; AAF1530.1; -.
DR HSSP: Q12923; 3PDZ.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 3.
DR SMART: SM00228; PDZ; 3.
DR PROSITE: PSS0106; PDZ; 3.
FT NON_TER 1
SQ SEQUENCE 1266 AA; 141071 MW; 92DF51B68081AA42 CRC64;
```

Query Match 36.9%; Score 500; DB 4; Length 1266;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 857 IADFKLTIVDDQKAGSRVGSGLGKSSSLSTLQTAFAEVTLNGDIPFHRPRRII 916
DB 767 IADFKLTIVDDQKAGSRVGSGLGKSSSLSTLQTAFAEVTLNGDIPFHRPRRII 826
QY 917 RGRGNESFRAIDKSYKPAVDDDDGEMTLIEDTESSSGRSVSTAADOSSHSLER 976
DB 917 RGRGNESFRAIDKSYKPAVDDDDGEMTLIEDTESSSGRSVSTAADOSSHSLER 976
```

```
DB 827 RGRGNESFRAIDKSYDKPAVDDDDGEMTLIEDTESSSRGRSVSTAADOSSHSLER 886
QY 977 QMNGOKEKDKTDKKDKTGKCKKDRDKEKDKAKKGMKGLCDMFRFGKHRDCKIE 1036
DB 887 QMNGOKEKDKTDKKDKTGKCKKDRDKEKDKAKKGMKGLCDMFRFGKHRDCKIE 946
QY 1037 KKGKIKIOESFTSEERIRMKOEERIOAKTREFREROAREVDVAEIDDFHRTGCCDEL 1096
DB 947 KKGKIKIOESFTSEERIRMKOEERIOAKTREFREROAREVDVAEIDDFHRTGCCDEL 1006
QY 1097 MTGVSVYEGSMALNARPOSREGHMDALYAQVKKPRNKRSPVDSNRSTPSNDRTOR 1156
DB 1007 MTGVSVYEGSMALNARPOSREGHMDALYAQVKKPRNKRSPVDSNRSTPSNDRTOR 1066
QY 1157 LKQEOQAKODEVDVDRRTYSFEQWPWPNARATQSGHSHSVVEYQMRQOREESSQO 1216
DB 1067 LKQEOQAKODEVDVDRRTYSFEQWPWPNARATQSGHSHSVVEYQMRQOREESSQO 1126
QY 1217 AQROYSLPQSRKNASSVSODSWEQNTSPGEGFQSAKENPYSYQSGRNGYLGHGFN 1276
DB 1127 AQROYSLPQSRKNASSVSODSWEQNTSPGEGFQSAKENPYSYQSGRNGYLGHGFN 1186
QY 1277 ARVMELELLRQEOQRREQOKKOPSESGPSNDYKRVQDPSYAPPKGFFRODVPSP 1336
DB 1187 ARVMELELLRQEOQRREQOKKOPSESGPSNDYKRVQDPSYAPPKGFFRODVPSP 1246
QY 1337 SOVARLNRLQTPKGRPPYS 1356
DB 1247 SOVARLNRLQTPKGRPPYS 1266
```

```
RESULT 6
Q9HC48 PRELIMINARY: PRT: 667 AA.
ID 09HC48:
AC 09HC48:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CTCL TUMOR ANTIGEN SE2-5 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=2143360; PubMed=11149944;
RA Eichmuller S., Usener D., Dummer R., Stein A., Thiel D.,
RA Schandendorf D.;
RT "Serological detection of cutaneous T-cell lymphoma-associated
RT antigens."
RL Proc. Natl. Acad. Sci. U S A. 98:629-634(2001).
DR EMBL: AF177228; AAG33676.1; -.
DR HSSP: Q12923; 3PDZ.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 2.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PSS0106; PDZ; 2.
FT NON_TER 1
SQ SEQUENCE 667 AA; 73499 MW; C653EC16802BAE02 CRC64;
```

Query Match 30.5%; Score 413; DB 4; Length 667;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 313 EHENLFRENDCIYRINDGLRNRRFEQAQHMFRQAMPTPIIWEHVPAANKQEOJLSOS 372
DB 1 EHENLFRENDCIYRINDGLRNRRFEQAQHMFRQAMPTPIIWEHVPAANKQEOJLSOS 60
QY 373 EKNNTYSSRFSPODYIDNRSVNSAGLHTVORARLHHPPEQIDSHSRLLPHSAHPSGKP 432
DB 373 EKNNTYSSRFSPODYIDNRSVNSAGLHTVORARLHHPPEQIDSHSRLLPHSAHPSGKP 432
```

```

Db 61 EKNNTSSRFSPDSQYIDNRVNSAGLHTVQAPRLNHPEQIDSHSLRPHSAHPSGKPP 120
OY 433 SAPASAPQNVFTTSSGVYTKIGKRLNIOLAKTEGLGFSITSRDVTIGSAPITYKN 492
Db 121 SAPASAPQNVFTTSSGVYTKIGKRLNIOLAKTEGLGFSITSRDVTIGSAPITYKN 180
OY 493 ILPRGAIDGRLKAGDRLIEVNGVDLYGKSQEEVVSLLRSTMEGTVSLVFRQEDAFH 552
Db 181 ILPRGAIDGRLKAGDRLIEVNGVDLYGKSQEEVVSLLRSTMEGTVSLVFRQEDAFH 240
OY 553 PRELNAEPSOMQIPRETKAEDEDIVLPDGTREFLFEVPLNDGSAAGVSKGNRSKE 612
Db 241 PRELNAEPSOMQIPRETKAEDEDIVLPDGTREFLFEVPLNDGSAAGVSKGNRSKE 300
OY 613 NHADLGFYKSIINGAASKDGRRLRVNDOLIAVNGESLLGKTQDMETLRRSMSTEGNK 672
Db 301 NHADLGFYKSIINGAASKDGRRLRVNDOLIAVNGESLLGKTQDMETLRRSMSTEGNK 360
OY 673 RGMIOILIVARRISKCNELKSPGSPGPPELPIETALDDRRRISHSLYSIGIEGDESFSRN 732
Db 361 RGMIOILIVARRISKCNELKSPGSPGPPELPIETALDDRRRISHSLYSIGIEGDESFSRN 420
OY 733 AALSRIMGSSGKYOLSPYVNMPODDVTIIEDDLPLVPLPHLSOOSSSSHDDVGFVTADA 792
Db 421 AALSRIMGSSGKYOLSPYVNMPODDVTIIEDDLPLVPLPHLSOOSSSSHDDVGFVTADA 480
OY 793 GTWAKAALSDSADCSLSPVDVPLAFQREGFGHQ 826
Db 481 GTWAKAALSDSADCSLSPVDVPLAFQREGFGHQ 514

```

```

RESULT 7
O96K28 ID 096K28 PRELIMINARY; PRT; 865 AA.
AC 096K28:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE CDNA FLJ14829 FIS, CLONE OVARC1000945, MODERATELY SIMILAR TO
DE RATUUS NORVEGICUS ATYPICAL PKC SPECIFIC BINDING PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-OVARIAN CARCINOMA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murekami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Niinomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project."
RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027735; BAB55330.1; -.
SQ SEQUENCE 865 AA; 95161 MW; 5378BBCD406D0835 CRC64;

```

Query Match 29.9%; Score 406; DB 4; Length 865;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 506; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 234 EKQEDDEDETEEDNSRVEPVGADTGLEHINPFSLDQNVKLVPEVNDGGLGIHVVPFSA 293
Db 99 EKQEDDEDETEEDNSRVEPVGADTGLEHINPFSLDQNVKLVPEVNDGGLGIHVVPFSA 158
OY 294 RGGRTVLGLLVKRLKKGKAEHNLFRFENDCIVARINDGDLNRRFEOAQHMFQAMRTPII 353
Db 159 RGGRTVLGLLVKRLKKGKAEHNLFRFENDCIVARINDGDLNRRFEOAQHMFQAMRTPII 218
OY 354 WFIHVPAANKQEYEQLSQSEKNNYSSRSPDSQYIDNRVNSAGLHTVQAPRLNHPE 413

```

```

Db 219 WFIHVPAANKQEYEQLSQSEKNNYSSRSPDSQYIDNRVNSAGLHTVQAPRLNHPE 278
OY 414 QIDSHSRRLPSHAPSGKPPASAPAPQNVFTTSSGVYTKIGKRLNIOLAKTEGLGF 473
Db 279 QIDSHSRRLPSHAPSGKPPASAPAPQNVFTTSSGVYTKIGKRLNIOLAKTEGLGF 338
OY 474 SITSRDVTIGSAPITYKNILPRGAIDGRLKAGDRLIEVNGVDLYGKSQEEVVSLLRS 533
Db 339 SITSRDVTIGSAPITYKNILPRGAIDGRLKAGDRLIEVNGVDLYGKSQEEVVSLLRS 398
OY 534 TKMEGTVSLVFRQEDAFHPRELNAEPSOMQIPRETKAEDEDIVLPDGTREFLFEVPL 593
Db 399 TKMEGTVSLVFRQEDAFHPRELNAEPSOMQIPRETKAEDEDIVLPDGTREFLFEVPL 458
OY 594 NDSGAGLVSVGNRKENHADIGIVKSTINGAASKDGRRLRVNDOLIAVNGESLLGK 653
Db 459 NDSGAGLVSVGNRKENHADIGIVKSTINGAASKDGRRLRVNDOLIAVNGESLLGK 518
OY 654 TNQDAMETLRRSMSTEGNKRGMIOLIVARRISKCNELKSPGSPGPPELPIETALDDRRR 713
Db 519 TNQDAMETLRRSMSTEGNKRGMIOLIVARRISKCNELKSPGSPGPPELPIETALDDRRR 578
OY 714 ISHSLYSIGIEGLDESPSRNALSRIMG 740
Db 579 ISHSLYSIGIEGLDESPSRNALSRIMG 605

```

```

RESULT 8
O9NWL4 ID 09NWL4 PRELIMINARY; PRT; 347 AA.
AC 09NWL4:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE CDNA FLJ20754 FIS, CLONE HEP02246 (UNKNOWN) (PROTEIN FOR
DE MGC:19518).
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawakami T., Noguchi S., Itoh T., Shigetani K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK00061; BAA91366.1; -.
DR EMBL; BC011711; AAH11711.1; -.
SQ SEQUENCE 347 AA; 40538 MW; BE2B3557996EC91E CRC64;

```

Query Match 25.6%; Score 347; DB 4; Length 347;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1010 MKKAKGMLKGLGDFRFGKRRKDDIKETGKIRKIOESFTSEERIRMKOERIOATRE 1069
Db 1 MKKAKGMLKGLGDFRFGKRRKDDIKETGKIRKIOESFTSEERIRMKOERIOATRE 60
OY 1070 FFRQARERDYAETIODFHRFFGCDDELMTYGVSSEGSMAINARPOSFREGHMDALYAO 1129
Db 61 FFRQARERDYAETIODFHRFFGCDDELMTYGVSSEGSMAINARPOSFREGHMDALYAO 120
OY 1130 VKKPRNKPSPVDSNRSTPSNHDRIQRLROFOAKODEVEDERRRYSFEOQWPNARPA 1189
Db 121 VKKPRNKPSPVDSNRSTPSNHDRIQRLROFOAKODEVEDERRRYSFEOQWPNARPA 180

```

QY 1190 TOSGRHSVSEVOMQROROEERESSOOAORVSLPROSCKNASSVSODSWEONYSPGEG 1249  
 DB 181 TOSGRHSVSEVOMQROROEERESSOOAORVSLPROSCKNASSVSODSWEONYSPGEG 240  
 QY 1250 FOSAKENPRYSYOGSNGSLGSGHNARVMELETOELLROBORKEOAKKOPPSGSPN 1309  
 DB 241 FOSAKENPRYSYOGSNGSLGSGHNARVMELETOELLROBORKEOAKKOPPSGSPN 300  
 QY 1310 YDSYKKVODPSYAPKGPGRDVPSPSOVARLNRLQTPKGRPFYS 1356  
 DB 301 YDSYKKVODPSYAPKGPGRDVPSPSOVARLNRLQTPKGRPFYS 347

RESULT 9  
 ID Q92340 PRELIMINARY; PRT; 1337 AA.  
 AC Q92340;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DE ATYPICAL PKC SPECIFIC BINDING PROTEIN.  
 GN ASP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98437350; PubMed=9763423;  
 RA Izumi Y., Hirose T., Tamai Y., Hirai S., Nagashima Y., Fujimoto T.,  
 RA Tabuse Y., Kempfues K.J., Ohno S.;  
 RT "An atypical PKC directly associates and colocalizes at the epithelial  
 RT tight junction with ASP, a mammalian homologue of caenorhabditis  
 RL U. Cell Biol. 143:95-106(1998).  
 DR EMBL; AB005549; BAA34216.1;  
 DR HSBP; Q12923; 3PDZ.  
 DR InterPro: IPR001478; PDZ.  
 DR Pfam: PF00595; PDZ.  
 DR SMART: SM00228; PDZ.  
 DR PROSITE: PS50106; PDZ.  
 SQ SEQUENCE 1337 AA; 149448 MW; EC980C5106B52F9C CRC64;

Query Match  
 Best Local Similarity 6.5%; Score 88; DB 11; Length 1337;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 566 PRETAEDEDIVLPDGTREFLFEVPLNDGSGAGVSVKGRSKENHADLGIFFVKSII 625  
 DB 566 PRETAEDEDIVLPDGTREFLFEVPLNDGSGAGVSVKGRSKENHADLGIFFVKSII 625  
 QY 626 NGGAASKDGRLRVNDOLIAVNGESLIGK 653  
 DB 626 NGGAASKDGRLRVNDOLIAVNGESLIGK 653  
 RESULT 10  
 ID Q99NH2 PRELIMINARY; PRT; 1333 AA.  
 AC Q99NH2;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DE PAR-3 180 KDA ISOFORM.  
 GN PAR3 OR PAR3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-NIH SWISS;  
 RX MEDLINE=99121117; PubMed=9920925;  
 RA Lin D., Gish G.D., Songyang Z., Pawson T.;  
 RT "The carboxyl terminus of B class ephrins constitutes a PDZ domain  
 RT binding motif".  
 RL J. Biol. Chem. 274:3726-3733(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NIH SWISS;  
 RX MEDLINE=20394297; PubMed=10934475;  
 RA Lin D., Edwards A.S., Fawcett J.P., Mbamalu G., Scott J.D., Pawson T.;  
 RT "A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and aPKC  
 RT signalling and cell polarity".  
 RL Nat. Cell Biol. 2:540-547(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NIH SWISS;  
 RA Lin D.C.J., Pawson T.J.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY026057; AAK07669.1;  
 DR HSBP; Q12923; 3PDZ.  
 DR MGI; 2135608; Pard3.  
 DR InterPro: IPR001478; PDZ.  
 DR Pfam: PF00595; PDZ.  
 DR SMART: SM00228; PDZ.  
 DR PROSITE: PS50106; PDZ.  
 SQ SEQUENCE 1333 AA; 149060 MW; AF67825C66DCFE86 CRC64;

Query Match  
 Best Local Similarity 5.7%; Score 77; DB 11; Length 1333;  
 Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 577 VLPDGTREFLFEVPLNDGSGAGVSVKGRSKENHADLGIFFVKSII 636  
 DB 577 VLPDGTREFLFEVPLNDGSGAGVSVKGRSKENHADLGIFFVKSII 636  
 QY 637 RVNDOLIAVNGESLIGK 653  
 DB 637 RVNDOLIAVNGESLIGK 653

RESULT 11  
 ID Q96DK9 PRELIMINARY; PRT; 624 AA.  
 AC Q96DK9;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DE CDNA FLJ25236 FIS, CLONE STM02096 (FRAGMENT).  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=GASTRIC MUCOSA;  
 RA Nimmiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 RA Furuya T., Takahashi M., Kikawa E., Omura Y., Abe K., Kamihara K.,  
 RA Nakagawa K., Sato K., Tanikawa M., Yamazaki M., Suzuki Y., Hata H.,  
 RA Irie K., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,  
 RA Nagai K., Isogai T., Sugano S.;  
 RT NEDO human cDNA sequencing project.  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK057965; BAB71623.1;  
 DR NON\_TER 624  
 FT 624  
 SQ SEQUENCE 624 AA; 68002 MW; 749EC5B79F9F096 CRC64;

Query Match  
 Best Local Similarity 1.4%; Score 19; DB 4; Length 624;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 882 LGKKSSLESLOTAAVEV 900  
 DB 462 LGKKSSLESLOTAAVEV 480

## RESULT 12

OY 096NX6 PRELIMINARY; PRT; 1143 AA.  
 AC 096NX6;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PARTITIONING-DEFECTIVE 3-LIKE PROTEIN SPLICING VARIANT B.  
 GN PAR3L.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gao L., Macara I.G., Joberty G.;  
 RT "Multiple splice variants of Par3 and of a novel related gene, Par3L,  
 RT produce functionally different proteins.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF428251; AAL3065.1; -;  
 SQ SEQUENCE 1143 AA; 126102 MW; 368B69CF81D45E7E CRC64;

## Query Match

Best Local Similarity 1.4%; Score 19; DB 4; Length 1143;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 882 LGKKSSLESLOTAAVEV 900  
 DB 677 LGKKSSLESLOTAAVEV 695

## RESULT 13

OY 096NX7 PRELIMINARY; PRT; 1205 AA.  
 AC 096NX7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PARTITIONING-DEFECTIVE 3-LIKE PROTEIN SPLICING VARIANT A.  
 GN PAR3L.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gao L., Macara I.G., Joberty G.;  
 RT "Multiple splice variants of Par3 and of a novel related gene, Par3L,  
 RT produce functionally different proteins.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF428250; AAL3066.1; -;  
 SQ SEQUENCE 1205 AA; 132494 MW; 26E704CCDCE8C8 CRC64;

## Query Match

Best Local Similarity 1.4%; Score 19; DB 4; Length 1205;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 882 LGKKSSLESLOTAAVEV 900  
 DB 739 LGKKSSLESLOTAAVEV 757

## RESULT 14

OY 096N09 PRELIMINARY; PRT; 545 AA.

AC 096N09;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CDNA FLJ31595 FIS, CLONE NT2R12002517, WEAKLY SIMILAR TO HOMO SAPIENS  
 DE PAR3 MRNA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nidomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,  
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yanashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahara K., Masuo Y., Nagai K., Isogai T.,  
 RT "NEO human cDNA sequencing project.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK056157; BAB71106.1; -;  
 SQ SEQUENCE 545 AA; 58691 MW; 03013B66B510FF0 CRC64;

## Query Match

Best Local Similarity 1.0%; Score 13; DB 4; Length 545;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 592 PLNDGSGAGLGV 604  
 DB 500 PLNDGSGAGLGV 512

## RESULT 15

OY 095LT2 PRELIMINARY; PRT; 128 AA.  
 AC 095LT2;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOPHETICAL 14.4 KDA PROTEIN.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 OC Cercopitheidae; Macaca.  
 OX NCBI\_Taxid=9341;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TESTIS;  
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,  
 RA Teruo K., Sugano S.;  
 RT "Isolation of novel full-length cDNA clones from macaque testis cDNA  
 RT libraries.";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB071110; BAB64504.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 128 AA; 14436 MW; AF92C43CC505F93F CRC64;

## Query Match

Best Local Similarity 0.9%; Score 12; DB 6; Length 128;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1325 KGFPRDVPSP 1336  
 DB 75 KGFPRDVPSP 86

Search completed: July 24, 2002, 14:37:02  
 Job time: 208 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:32:04 ; Search time 18.11 Seconds  
(without alignments)  
1828.885 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 1356  
Sequence: 1 MKVTVCGRTRRVVPCGDGH.....SQVARLNRLQTPKGRPEYS 1356

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 231628 segs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8	0.6	31	3	US-09-045-632-70	Sequence 70, Appl
2	8	0.6	31	3	US-09-045-632-73	Sequence 73, Appl
3	8	0.6	79	3	US-09-100-804-26	Sequence 26, Appl
4	8	0.6	297	2	US-09-151-611-3	Sequence 3, Appl1
5	8	0.6	297	4	US-09-370-102-3	Sequence 3, Appl1
6	8	0.6	348	3	US-08-415-655-5	Sequence 5, Appl1
7	8	0.6	348	3	US-08-415-655-13	Sequence 13, Appl
8	8	0.6	348	3	US-08-415-655-15	Sequence 15, Appl
9	8	0.6	388	1	US-08-445-640-6	Sequence 6, Appl1
10	8	0.6	388	3	US-08-170-558-6	Sequence 6, Appl1
11	8	0.6	388	3	US-08-447-314-6	Sequence 6, Appl1
12	8	0.6	388	3	US-08-445-461-6	Sequence 6, Appl1
13	8	0.6	450	2	US-08-665-037-2	Sequence 2, Appl1
14	8	0.6	450	2	US-08-666-067-2	Sequence 2, Appl1
15	8	0.6	450	2	US-08-732-870-2	Sequence 2, Appl1
16	8	0.6	890	1	US-08-445-640-2	Sequence 2, Appl1
17	8	0.6	890	3	US-08-170-558-2	Sequence 2, Appl1
18	8	0.6	890	3	US-08-447-314-2	Sequence 2, Appl1
19	8	0.6	890	3	US-08-445-461-2	Sequence 2, Appl1
20	8	0.6	911	1	US-08-286-305A-1	Sequence 1, Appl1
21	8	0.6	911	2	US-08-441-104A-1	Sequence 1, Appl1
22	8	0.6	911	2	US-08-440-816A-1	Sequence 1, Appl1
23	8	0.6	911	4	US-09-417-381A-1	Sequence 1, Appl1
24	8	0.5	61	4	US-09-314-268-145	Sequence 145, App
25	8	0.5	100	3	US-08-851-843A-10	Sequence 10, Appl
26	7	0.5	100	4	US-08-974-549A-192	Sequence 192, App
27	7	0.5	100	4	US-08-854-050-10	Sequence 10, Appl

28	7	0.5	100	4	US-09-430-323-10	Sequence 10, Appl
29	7	0.5	128	4	US-09-199-637A-79	Sequence 79, Appl
30	7	0.5	182	1	US-08-345-756-2	Sequence 2, Appl1
31	7	0.5	182	1	US-08-625-198-2	Sequence 2, Appl1
32	7	0.5	305	3	US-08-965-600-1	Sequence 1, Appl1
33	7	0.5	320	2	US-08-800-264A-6	Sequence 6, Appl1
34	7	0.5	320	2	US-09-018-628-6	Sequence 6, Appl1
35	7	0.5	320	3	US-09-273-378-6	Sequence 6, Appl1
36	7	0.5	320	3	US-09-209-605-6	Sequence 6, Appl1
37	7	0.5	353	3	US-08-966-318-3	Sequence 3, Appl1
38	7	0.5	353	4	US-09-216-619-3	Sequence 3, Appl1
39	7	0.5	377	3	US-08-888-077A-27	Sequence 27, Appl
40	7	0.5	384	3	US-08-946-026-27	Sequence 27, Appl
41	7	0.5	437	2	US-08-800-264A-7	Sequence 7, Appl1
42	7	0.5	437	2	US-09-018-628-7	Sequence 7, Appl1
43	7	0.5	437	3	US-09-273-378-7	Sequence 7, Appl1
44	7	0.5	437	3	US-09-209-605-7	Sequence 7, Appl1
45	7	0.5	439	1	US-08-519-103-12	Sequence 12, Appl

#### ALIGNMENTS

RESULT 1  
US-09-045-632-70  
; Sequence 70, Application US/09045632  
; Patent No. 6001575  
; GENERAL INFORMATION:  
; APPLICANT: HUGANIT, Richard L.  
; APPLICANT: Dong, Hualing  
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND  
; TITLE OF INVENTION: GRIP-RELATED MOLECULES  
; NUMBER OF SEQUENCES: 105  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/045,632  
; FILING DATE: 19-MAR-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/041,016  
; FILING DATE: 19-MAR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cortless, Peter F.  
; REGISTRATION NUMBER: 33,860  
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-523-3400  
; TELEFAX: 617-523-6440  
; INFORMATION FOR SEQ ID NO: 70:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 31 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-09-045-632-70

Query Match 0.6%; Score 8; DB 3; Length 31;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 633 DGLRVND 640  
Db 15 DGLRVND 22

## RESULT 2

US-09-045-632-73  
; Sequence 73, Application US/09045632  
; Patent No. 6001575  
; GENERAL INFORMATION:  
; APPLICANT: HUGANIR, Richard L.  
; APPLICANT: Dong, Hualing  
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND  
; TITLE OF INVENTION: GRIP-RELATED MOLECULES  
; NUMBER OF SEQUENCES: 105  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/045,632  
; FILING DATE: 19-MAR-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/041,016  
; FILING DATE: 19-MAR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Corless, Peter F.  
; REGISTRATION NUMBER: 33,860  
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-523-3400  
; TELEFAX: 617-523-6440  
; INFORMATION FOR SEQ ID NO: 73:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 31 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-09-045-632-73

Query Match 0.6%; Score 8; DB 3; Length 31;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 633 DGLRVND 640  
Db 15 DGLRVND 22

## RESULT 3

US-09-100-804-26  
; Sequence 26, Application US/09100804  
; Patent No. 6066472  
; GENERAL INFORMATION:  
; APPLICANT: GONEZ, LEONEL JORGE  
; APPLICANT: SARAS, JAN  
; APPLICANT: CLAESSEN-WELSH, LENA  
; APPLICANT: HELDIN, CARL-HENRIK  
; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
; TITLE OF INVENTION: TYROSINE PHOSPHATASES  
; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
; STREET: 600 ATLANTIC AVENUE  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/100,804  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/596,291  
; FILING DATE: 09-AUG-1996  
; APPLICATION NUMBER: US 08/115,573  
; FILING DATE: 01-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/09943  
; FILING DATE: 01-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GATES, EDWARD R.  
; REGISTRATION NUMBER: 31,616  
; REFERENCE/DOCKET NUMBER: 10461/7003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 79 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; US-09-100-804-26

Query Match 0.6%; Score 8; DB 3; Length 79;  
Best Local Similarity 100.0%; Pred. No. 5.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 633 DGLRVND 640  
Db 39 DGLRVND 46

## RESULT 4

US-09-151-611-3  
; Sequence 3, Application US/09151611  
; Patent No. 5958731  
; GENERAL INFORMATION:  
; APPLICANT: Yue, Henry  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Patterson, Chandra  
; TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN  
; FILE REFERENCE: PR-0599 US  
; CURRENT APPLICATION NUMBER: US/09/151,611  
; CURRENT FILING DATE: 1998-09-11  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PERL Program  
; SEQ ID NO 3  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE: -  
; OTHER INFORMATION: g1685067



US-09-151-611-3

Query Match 0.6%; Score 8; DB 2; Length 297;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 640 DOLIAVNG 647  
|||||  
Db 245 DOLIAVNG 252

RESULT 5  
US-09-370-102-3  
; Sequence 3, Application US/09370102  
; Patent No. 6265547  
; GENERAL INFORMATION:  
; APPLICANT: Yue, Henry  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Patterson, Chandra  
; TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN  
; FILE REFERENCE: PF-0599 US  
; CURRENT APPLICATION NUMBER: US/09/370,102  
; EARLIER APPLICATION NUMBER: 09/151,611  
; EARLIER FILING DATE: 1998-09-11  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PERL Program  
; SEQ ID NO 3  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE: -  
; OTHER INFORMATION: g1685067  
US-09-370-102-3

Query Match 0.6%; Score 8; DB 4; Length 297;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 640 DOLIAVNG 647  
|||||  
Db 245 DOLIAVNG 252

RESULT 6  
US-08-415-655-5  
; Sequence 5, Application US/08415655  
; Patent No. 6025480  
; GENERAL INFORMATION:  
; APPLICANT: Massague, Joan  
; APPLICANT: Lee, Mong-hong  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: P75KIP2, A CYCLIN-DEPENDENT KINASE INHIBITOR AND USES OF  
; TITLE OF INVENTION: SAME  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/415,655  
; FILING DATE:  
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 1747/47418  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 348 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-415-655-5

Query Match 0.6%; Score 8; DB 3; Length 348;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 431 PPSAPASA 438  
|||||  
Db 139 PPSAPASA 146

RESULT 7  
US-08-415-655-13  
; Sequence 13, Application US/08415655  
; Patent No. 6025480  
; GENERAL INFORMATION:  
; APPLICANT: Massague, Joan  
; APPLICANT: Lee, Mong-hong  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: P75KIP2, A CYCLIN-DEPENDENT KINASE INHIBITOR AND USES OF  
; TITLE OF INVENTION: SAME  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/415,655  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 1747/47418  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 348 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-415-655-13

Thu Jul 25 08:38:17 2002

Query Match 0.6%; Score 8; DB 3; Length 348;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 431 PPSAPASA 438  
DB 139 PPSAPASA 146

RESULT 8  
US-08-415-655-15  
Sequence 15, Application US/08415655  
Patent No. 6025480  
GENERAL INFORMATION:  
APPLICANT: Massague, Joan  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING  
TITLE OF INVENTION: P75KIP2, A CYCLIN-DEPENDENT KINASE INHIBITOR AND USES OF  
TITLE OF INVENTION: SAME  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/08/415,655  
APPLICATION NUMBER: US/08-415,655  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 1747/47418  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ. ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 348 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-415-655-15

Query Match 0.6%; Score 8; DB 3; Length 348;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 431 PPSAPASA 438  
DB 139 PPSAPASA 146

RESULT 9  
US-08-445-640-6  
Sequence 6, Application US/08445640  
Patent No. 5709858  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases

NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION NUMBER: US/08/445,640  
APPLICATION NUMBER: US/08-445,640  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ. ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-445-640-6

Query Match 0.6%; Score 8; DB 1; Length 388;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 434 APASAPON 441  
DB 281 APASAPON 288

RESULT 10  
US-08-170-558-6  
Sequence 6, Application US/08170558  
Patent No. 6001621  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/170,558  
FILING DATE: 20-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-170-558-6

Query Match  
Best Local Similarity 100.0%; Score 8; DB 3; Length 388;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 APASAPON 441  
DB 281 APASAPON 288

RESULT 11  
US-08-447-314-6  
Sequence 6, Application US/08447314  
Patent No. 6087144  
GENERAL INFORMATION:  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,314  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C1D2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:

LENGTH: 388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-447-314-6

Query Match  
Best Local Similarity 100.0%; Score 8; DB 3; Length 388;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 APASAPON 441  
DB 281 APASAPON 288

RESULT 12  
US-08-445-461-6  
Sequence 6, Application US/08445461  
Patent No. 6096527  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,461  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-445-461-6

Query Match  
Best Local Similarity 100.0%; Score 8; DB 3; Length 388;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 APASAPON 441  
DB 281 APASAPON 288

## RESULT 13

US-08-665-037-2  
Sequence 2, Application US/08665037  
Patent No. 5895813  
GENERAL INFORMATION:  
APPLICANT: Seedorf, Klaus  
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT  
OF TKA-1 RELATED  
TITLE OF INVENTION: OF TKA-1 RELATED  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/665,037  
FILING DATE: June 13, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/005,167  
FILING DATE: October 13, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 220/156  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 450 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-665-037-2

Query Match 0.6%; Score 8; DB 2; Length 450;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 509 DRLIEVNG 516  
|||||||  
Db 193 DRLIEVNG 200

## RESULT 14

US-08-666-067-2  
Sequence 2, Application US/08666067  
Patent No. 5922842  
GENERAL INFORMATION:  
APPLICANT: Seedorf, Klaus  
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT  
OF TKA-1 RELATED  
TITLE OF INVENTION: OF TKA-1 RELATED  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street

STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/666,067  
FILING DATE: June 13, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/005,421  
FILING DATE: October 13, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 220/157  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 450 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-666-067-2

Query Match 0.6%; Score 8; DB 2; Length 450;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 509 DRLIEVNG 516  
|||||||  
Db 193 DRLIEVNG 200

## RESULT 15

US-08-732-870-2  
Sequence 2, Application US/08732870  
Patent No. 5945523  
GENERAL INFORMATION:  
APPLICANT: Seedorf, Iulgard  
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT  
OF TKA-1 RELATED  
TITLE OF INVENTION: OF TKA-1 RELATED  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/732,870  
FILING DATE: October 15, 1996  
CLASSIFICATION: 536

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/666,037  
;; FILING DATE: June 13, 1996  
;; APPLICATION NUMBER: 08/666,067  
;; FILING DATE: June 13, 1996  
;; APPLICATION NUMBER: 60/005,167  
;; FILING DATE: October 13, 1995  
;; APPLICATION NUMBER: 60/005,421  
;; FILING DATE: October 13, 1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Warburg, Richard J.  
;; REGISTRATION NUMBER: 32,327  
;; REFERENCE/DOCKET NUMBER: 222/247  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (213) 489-1600  
;; TELEFAX: (213) 955-0440  
;; TELEX: 67-3510  
;; INFORMATION FOR SEQ. ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 450 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-732-870-2

Query Match 0.6%; Score 8; DB 2; Length 450;  
Best local Similarity 100.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 509 DRLIEVNG 516  
|||  
Db 193 DRLIEVNG 200

Search completed: July 24, 2002, 14:35:23  
Job time: 199 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:34:09 ; Search time 16.09 Seconds

(without alignments)  
3263.127 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 1356  
Sequence: 1 MKYTVCFGRTRVVVPCGDGH.....SQVARNLRLOTPEKGRPFYS 1356

Scoring table: OLIGO  
Gapop 60.0 , Capext 60.0

Searched: 105224 seqs, 38719550 residues

Word size: 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	0.7	1693	1 Y163_SYNY3	05563 synechocyst
2	8	0.6	200	1 PAD_BACFI	P94300 bacillus fi
3	8	0.6	283	1 P635_METUA	Q58052 methanococ
4	8	0.6	348	1 CDNC_MOUSE	P49919 mus musculu
5	8	0.6	475	1 Z131_HUMAN	P52739 homo sapien
6	8	0.6	488	1 VAB1_HORVU	Q40078 hordeum vul
7	8	0.6	592	1 YK06_CAEEL	P34294 caenorhabdi
8	8	0.6	724	1 DLG4_MOUSE	Q62108 mus musculu
9	8	0.6	724	1 DLG4_RAT	P31016 rattus norv
10	8	0.6	737	1 YNC2_CAEEL	P34835 caenorhabdi
11	8	0.6	767	1 DLG4_HUMAN	P78352 homo sapien
12	8	0.6	852	1 DLG2_RAT	Q63622 rattus norv
13	8	0.6	870	1 DLG2_HUMAN	Q15700 homo sapien
14	8	0.6	890	1 TYO3_HUMAN	Q06418 homo sapien
15	8	0.6	904	1 DLG1_HUMAN	Q12959 homo sapien
16	8	0.6	911	1 DLG1_RAT	Q62996 rattus norv
17	8	0.6	1081	1 SY1_TERTH	P36422 tetrahymena
18	8	0.6	1490	1 CRK7_HUMAN	Q9nyv4 homo sapien
19	8	0.6	4036	1 RRP1_DUGBU	Q66431 dupre virus
20	7	0.5	79	1 Y476_ARCFU	Q29774 archaeoglob
21	7	0.5	107	1 KVID_HUMAN	P01596 homo sapien
22	7	0.5	116	1 Y789_METUA	Q58199 methanococ
23	7	0.5	142	1 YOHN_BACSU	P54512 bacillus su
24	7	0.5	150	1 VE6_HPV42	P27229 human papil
25	7	0.5	177	1 RL10_MYCLE	Q9cbk7 mycobacteri
26	7	0.5	178	1 RL10_MYCLE	P96940 mycobacteri
27	7	0.5	180	1 YRBI_HAEIN	P45314 haemophilus
28	7	0.5	181	1 OM20_NEUCR	P35848 neuropsora
29	7	0.5	246	1 YIAT_ECO57	P58242 escherichia
30	7	0.5	246	1 YIAT_ECOLI	P37681 escherichia
31	7	0.5	258	1 TPM3_CHICK	P19353 gallus gall
32	7	0.5	248	1 YNM8_YEAST	P53862 saccharomyc
33	7	0.5	261	1 MOTB_BACSU	P28612 bacillus su

34	7	0.5	270	1 VA32_VARV	P33849 variola vir
35	7	0.5	276	1 ATP6_SYNY3	P27178 synechocyst
36	7	0.5	287	1 T215_ARATH	O23403 arabidopsis
37	7	0.5	297	1 YNFI_ECOLI	P75559 escherichia
38	7	0.5	300	1 VA32_VACCV	P21055 vaccinia vl
39	7	0.5	312	1 BLAB_STRCI	P33652 streptomyc
40	7	0.5	342	1 LP1X_CLOPE	P18020 clostridium
41	7	0.5	354	1 LPXD_CHLMU	Q9pkf1 chlamydia m
42	7	0.5	354	1 RECA_SYNY3	P74737 synechocyst
43	7	0.5	371	1 OP25_HAEIN	P46027 haemophilus
44	7	0.5	376	1 PSD4_MOUSE	Q35226 mus musculu
45	7	0.5	377	1 NTV1_ANANP	Q44290 anabaena sp

## ALIGNMENTS

```
RESULT 1
Y163_SYNY3 STANDARD; PRT; 1693 AA.
AC 05563;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical WD-repeat protein SL0163.
GN SL0163.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 644 to 924 of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- SIMILARITY: CONTAINS 16 WD REPEATS (TRP-ASP DOMAINS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: D63999; BAA10064.1; -.
CC InterPro: IPR001680; WD40.
CC Pfam: PF00400; WD40; 16.
CC PRINTS: PR00320; GPOTREINRPT.
CC SMART: SMD0320; WD40; 16.
CC PROSITE: PS00678; WD_REPEATS_1; 8.
CC PROSITE: PS50082; WD_REPEATS_2; 15.
CC PROSITE: PS50294; WD_REPEATS_REGION; 1.
CC Hypothetical protein; Repeat; WD repeat; Complete proteome.
CC REPEAT 1008 1042 WD 1.
CC REPEAT 1053 1083 WD 2.
CC REPEAT 1094 1124 WD 3.
CC REPEAT 1135 1165 WD 4.
CC REPEAT 1176 1206 WD 5.
CC REPEAT 1217 1247 WD 6.
CC REPEAT 1258 1288 WD 7.
CC REPEAT 1299 1329 WD 8.
CC REPEAT 1340 1370 WD 9.
CC REPEAT 1381 1411 WD 10.
CC REPEAT 1422 1452 WD 11.
CC REPEAT 1463 1493 WD 12.
CC REPEAT 1504 1534 WD 13.
CC REPEAT 1545 1575 WD 14.
CC REPEAT 1586 1616 WD 15.
CC REPEAT 1627 1657 WD 16.
```

SO SEQUENCE 1693 AA; 189935 MW; 0977A827A0251CFP CRC64;

Query Match  
Best Local Similarity 100.0%; Score 9; DB 1; Length 1693;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 381 RFPSPDSQYI 389  
DB 1431 RFPSPDSQYI 1439

#### RESULT 2

PAAD\_BACFI STANDARD; PRT; 200 AA.

AC P94300;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
OS Probable aromatic acid decarboxylase (EC 4.1.1.-).  
OC Bacillus firmus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1399;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-OF4;  
RA Ito M., Kuriwaki T.A.;  
RT Cloning and sequence of gerc locus from alkaliphilic Bacillus firmus OF4.  
RT Submitted (JUN-1996) to the EMBL/Genbank/DDJJ databases.  
CC -1- SIMILARITY: BELONGS TO THE POLYPEPTYL P-HYDROXYBENZOATE / PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U61168; AAB41845.1; -  
DR InterPro; IPR003382; Flavoprotein.  
DR Pfam; PF02441; Flavoprotein; 1.  
KW Hypothetical protein; lyase; Decarboxylase.  
SQ SEQUENCE 200 AA; 22233 MW; 0C212E8AD11AC13D CRC64;

Query Match  
Best Local Similarity 100.0%; Score 8; DB 1; Length 200;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1283 TOELLROE 1290  
DB 26 TOELLROE 33

#### RESULT 3

X635\_METHA STANDARD; PRT; 283 AA.

AC Q58052;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein M0655.  
GN M0655.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
OC Methanococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fumman J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii."  
RL Science 273:1058-1073(1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U67511; AAB98633.1; -  
DR TIGR; M0635;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 283 AA; 32502 MW; 1CE52FC8457C5E20 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 8; DB 1; Length 283;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 999 KKKDRDKE 1006  
DB 245 KKKDRDKE 252

#### RESULT 4

CDNC\_MOUSE STANDARD; PRT; 348 AA.

AC P49919;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Cyclin-dependent kinase inhibitor 1C (cyclin-dependent kinase inhibitor p57) (P57KIP2).  
DE CDKN1C OR KIP2.  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=95247027; PubMed=7729683;  
RA Lee M.-H., Reynolds I., Massague J.;  
RT "Cloning of p57KIP2, a cyclin-dependent kinase inhibitor with unique domain structure and tissue distribution."  
RL Genes Dev. 9:639-649(1995).  
RN [2]  
RP SEQUENCE OF 14-348 FROM N.A.  
RX MEDLINE=95247028; PubMed=7729684;  
RA Matsuno S., Edwards M.C., Bai C., Parker S., Zhang P., Baldini A., Harper J.W., Elledge S.J.;  
RT "p57KIP2, a structurally distinct member of the p21CIP1 Cdk inhibitor family, is a candidate tumor suppressor gene."  
RL Genes Dev. 9:650-662(1995).  
CC -1- FUNCTION: POTENT TIGHT-BINDING INHIBITOR OF SEVERAL G1 CYCLIN/CDK COMPLEXES (CYCLIN E-CDK2, CYCLIN D2-CDK4, AND CYCLIN A-CDK2) AND, TO LESSER EXTENT, OF THE MITOTIC CYCLIN B-CDK2. NEGATIVE REGULATOR OF CELL PROLIFERATION. MAY PLAY A ROLE IN MAINTENANCE OF THE NONPROLIFERATIVE STATE THROUGHOUT LIFE.  
CC -1- SUBCELLULAR LOCATION: Nuclear.



CC -1- TISSUE SPECIFICITY: IT IS EXPRESSED IN THE HEART, BRAIN, LUNG,  
 CC SKELETAL MUSCLE, KIDNEY, PANCREAS AND TESTIS. HIGH LEVELS ARE SEEN  
 CC IN THE PLACENTA WHILE LOW LEVELS ARE SEEN IN THE LIVER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U20553; AAC52186.1; -  
 CC EMBL: U22399; AAA85096.1; -  
 CC HSSP: P46527; IJ5U.  
 CC MGD: MGI:104564; Cdkn1c.  
 CC InterPro: IPR003175; CDI.  
 CC Pfam: PF02234; CDI.1.  
 CC Cell cycle; Alternative splicing.  
 CC FT DOMAIN 108 189 PRO-RICH.  
 CC FT DOMAIN 178 284 GLU/ASP-RICH.  
 CC FT DOMAIN 309 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 CC FT VARSPPLIC 1 13 MISSING (IN ISOFORM KIP2B/P57B).  
 CC FT CONFLICT 150 151 DA -> EP (IN REF. 2).  
 CC SQ SEQUENCE 348 AA: 37331 MW: 108A8538D77016D9 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 431 PPSAPASA 438  
 |||||||  
 Db 139 PPSAPASA 146

DR MIM: 604073; -  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR Pfam: PF00096; zf-C2H2; 5.  
 DR SMART: SM00355; ZNF\_C2H2; 5.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
 DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 5.  
 DR Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;  
 KW Nuclear protein; Repeat.  
 FT NON\_TER 1 1  
 FT DOMAIN 113 295 ZINC FINGERS.  
 FT ZN\_FING 113 135 C2H2-TYPE.  
 FT ZN\_FING 140 163 C2H2-TYPE.  
 FT ZN\_FING 180 202 C2H2-TYPE.  
 FT ZN\_FING 244 266 C2H2-TYPE.  
 FT ZN\_FING 272 295 C2H2-TYPE.  
 SQ SEQUENCE 475 AA: 54087 MW: 3388B15B8CC567FF CRC64;

Query Match 0.6%; Score 8; DB 1; Length 475;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1208 QERRESSQ 1215  
 |||||||  
 Db 426 QERRESSQ 433

RESULT 5  
 Z131\_HUMAN STANDARD; PRT; 475 AA.  
 AC P52739;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Zinc finger protein 131 (Fragment).  
 GN ZNF131.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Insulinoma;  
 RX MEDLINE=96044430; PubMed=7557990;  
 RA Tomerup N., Vissing H.;  
 RT Isolation and fine mapping of 16 novel human zinc finger-encoding  
 RT cDNAs identify putative candidate genes for developmental and  
 RT malignant disorders.";  
 RL Genomics 27:259-264(1995).  
 CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U09410; AAC50251.1; -  
 CC HSSP: P07248; ZADR.

RESULT 6  
 VAB1\_HORVU STANDARD; PRT; 488 AA.  
 ID VAB1\_HORVU  
 AC Q40078;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Vacuolar ATP synthase subunit B isoform 1 (EC 3.6.3.14) (V-ATPase B  
 DE subunit 1) (Vacuolar proton pump B subunit 1).  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Root;  
 RX MEDLINE=94159793; PubMed=8115549;  
 RA Berkelman T., Houtchens K.A., Dupont F.M.;  
 RT "Two cDNA clones encoding isoforms of the B subunit of the vacuolar  
 RT ATPase from barley roots.";  
 RL Plant Physiol. 104:287-288(1994).  
 CC -1- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF  
 CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY  
 CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +  
 CC H(+) (out).  
 CC -1- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A  
 CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,  
 CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE  
 CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).  
 CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L11862; AAA81330.1; -  
 CC InterPro: IPR004100; ATP-synt\_ab\_N.  
 CC InterPro: IPR000793; ATPase\_AB\_C.  
 CC InterPro: IPR000194; ATPase\_alpha\_beta.  
 CC Pfam: PF00006; ATP-synt\_ab; 1.

DR Pfam: PF00306; ATP-synt\_ab\_C: 1.  
DR Pfam: PF02874; ATP-synt\_ab\_N: 1.  
DR PROSITE: PS00152; ATPASE\_ALPHA\_BETA: 1.  
KW Hydrolyase; ATP synthetase; Hydrogen ion transport; Multigene family.  
SQ SEQUENCE 488 AA; 54026 MW; 0571B898CECC1070 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 488;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 302 LVKRLKRG 309  
Db 186 LVKRLKRG 193

RESULT 7  
YK06.CAEEL STANDARD; PRT; 592 AA.  
AC P34294;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Hypothetical 66.9 kDa protein C05B5.6 in Chromosome III.  
GN C05B5.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxId=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Morimore B.J.;  
RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: SOME. TO C.ELEGANS ZK1290.9.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: Z32679; GAA83594.1; -  
DR PIR: S43570; S43570.  
DR WormPep: C05B5.6; CE00052.  
DR InterPro: IPR002900; DUF38.  
DR Pfam: PF01827; DUF38; 1.  
KW Hypothetical protein.  
FT DOMAIN 9 15 POLY-SER.  
FT DOMAIN 448 455 POLY-LEU.  
FT DOMAIN 584 587 POLY-SER.  
SQ SEQUENCE 592 AA; 66879 MW; AOBCA915649710BD CRC64;

Query Match 0.6%; Score 8; DB 1; Length 592;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 774 SDOSSSSS 781  
Db 6 SDOSSSSS 13

RESULT 8  
DLAG\_MOUSE STANDARD; PRT; 724 AA.  
ID DLAG\_MOUSE  
AC Q62108;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)

DE (Synapse-associated protein 90) (Discs, large homolog 4).  
GN DLAG OR DLAG4 OR PSD95.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-DRA/2; TISSUE-Brain;  
RA Kohmura N., Yagi T.;  
RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR  
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY  
CC ON THE PRESYNAPTIC SIDE (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 3 PDZ/DRH DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: D50621; BAA09297.1; -  
DR HSSP: P31016; IBE9.  
DR MGD: MGI:1277959; DLAG4.  
DR InterPro: IPR000619; Guanylate\_kin.  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00625; Guanylate\_kin; 1.  
DR Pfam: PF00595; PDZ; 3.  
DR Pfam: PF00018; SH3; 1.  
DR SMART: SM00072; Gukc; 1.  
DR SMART: SM00228; PDZ; 3.  
DR SMART: SM00326; SH3; 1.  
DR PROSITE: PS00856; GUANYLATE\_KINASE\_1; 1.  
DR PROSITE: PS50052; GUANYLATE\_KINASE\_2; 1.  
DR PROSITE: PS50106; PDZ; 3.  
DR PROSITE: PS50002; SH3; 1.  
KW SH3 domain; Repeat.  
FT DOMAIN 65 151 PDZ 1.  
FT DOMAIN 160 246 PDZ 2.  
FT DOMAIN 313 393 PDZ 3.  
FT DOMAIN 428 498 SH3.  
FT DOMAIN 534 724 GUANYLATE\_KINASE.  
SQ SEQUENCE 724 AA; 80472 MW; 7EFC99E1FF9DBA CRC64;

Query Match 0.6%; Score 8; DB 1; Length 724;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 633 DGRLRVND 640  
Db 108 DGRLRVND 115

RESULT 9  
DLAG\_RAT STANDARD; PRT; 724 AA.  
ID DLAG\_RAT  
AC P31016; P97631;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)  
DE (Synapse-associated protein 90) (Discs, large homolog 4).  
GN DLAG OR PSD95.  
OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Brain;  
 RX MEDLINE=93040233; PubMed=1419001;  
 RA Cho K.-O., Hunt C.A., Kennedy M.B.;  
 RT "The rat brain postsynaptic density fraction contains a homolog of  
 RT the Drosophila discs-large tumor suppressor protein.";  
 RL Neuron 9:929-942(1992).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Brain;  
 RX MEDLINE=93186749; PubMed=7680343;  
 RA Kistner U., Wenzel B.M., Voh R.W., Cases-Ianghoff C., Garner A.M.,  
 RA Appeltner U., Voss B., Gundelfinger E.D., Garner C.C.;  
 RT "SNAP90, a rat presynaptic protein related to the product of the  
 RT Drosophila tumor suppressor gene d19-A.";  
 RL J. Biol. Chem. 268:4580-4583(1993).  
 RN [3]  
 RN SEQUENCE OF 566-625 FROM N.A.  
 RC STRAIN-WISTAR KYOTO; TISSUE=Vascular smooth muscle;  
 RA Adams L.D., Meryn I., Schwartz S.M.;  
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RN X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS) OF 302-402.  
 RX MEDLINE=96270509; PubMed=8674113;  
 RA Doyle D.A., Lee A., Lewis J., Kim E., Sheng M., Mackinnon R.;  
 RT "Crystal structures of a complexed and peptide-free membrane protein-  
 RT binding domain: molecular basis of peptide recognition by PDZ.";  
 RL Cell 85:1067-1076(1996).  
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR  
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY  
 CC ON THE PRESYNAPTIC SIDE (WAS ORIGINALLY THOUGHT TO BE  
 CC POSTSYNAPTIC).  
 CC -1- TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.  
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DRH DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M96853; AAA41971.1; -  
 DR EMBL: X66474; CAA47103.1; -  
 DR EMBL: U77090; AAB38270.1; -  
 DR PIR: S26407; S26407.  
 DR PIR: JH0800; JH0800.  
 DR PDB: 1BE9; 21-OCY-98.  
 DR PDB: 1BE9; 21-OCY-98.  
 DR InterPro: IPR000619; Guanylate\_kin.  
 DR InterPro: IPR001478; PDZ.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00625; Guanylate\_kin. 1.  
 DR Pfam: PF00018; SH3. 1.  
 DR SMART: SM00072; GUKC. 1.  
 DR SMART: SM00228; PDZ. 3.  
 DR SMART: SM00326; SH3. 1.  
 DR PROSITE: PS00856; GUANYLATE\_KINASE\_1; 1.  
 DR PROSITE: PS50052; GUANYLATE\_KINASE\_2; 1.  
 DR PROSITE: PS50106; PDZ. 3.  
 DR PROSITE: PS50002; SH3. 1.  
 DR SH3 domain; Repeat; 3D-structure.  
 FT DOMAIN 65 151 PDZ 1.

FT DOMAIN 160 246 PDZ 2.  
 FT DOMAIN 313 393 PDZ 3.  
 FT DOMAIN 428 498 SH3.  
 FT DOMAIN 534 724 GUANYLATE KINASE.  
 FT CONFLICT 61 61 M -> L (IN REF. 2).  
 FT CONFLICT 78 78 S -> T (IN REF. 2).  
 FT CONFLICT 177 182 GWGNH -> ALGST (IN REF. 2).  
 FT CONFLICT 200 200 A -> G (IN REF. 2).  
 FT CONFLICT 254 254 S -> T (IN REF. 2).  
 FT CONFLICT 539 555 ILGPRKDRANDDLSEF -> ISLPKRTVPMISPPS  
 FT CONFLICT 623 625 (IN REF. 2).  
 FT CONFLICT 625 625 (IN REF. 3).  
 FT CONFLICT 724 AA; 80465 MW; 7922D4E8E0F9AD85 CRC64;  
 SQ SEQUENCE  
 Query Match 0.6%; Score 8; DB 1; Length 724;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 633 DGRLRVND 640  
 Db 108 DGRLRVND 115  
 RESULT 10  
 YNC2\_CAEEL  
 ID YNC2\_CAEEL STANDARD; PRT; 737 AA.  
 AC P34535;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 83.6 kDa protein R05D3.2 in chromosome III.  
 GN R05D3.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berts M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Crawford M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,  
 RA Fulton L., Garner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Showkeen R.,  
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Wohlschlag P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: I07144; AK21441.1; -  
 DR PIR: S44862; S44862.  
 DR WormPep: R05D3.2; CE00281.  
 KW Hypothetical protein.  
 SQ SEQUENCE 737 AA; 83555 MW; 3397543C5C6EC9B4 CRC64;  
 Query Match 0.6%; Score 8; DB 1; Length 737;  
 Best Local Similarity 100.0%; Pred. No. 17;

Thu Jul 25 08:38:25 2002

```

Matches      8;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

QY      376  NYSRSRS 383
         |||||||
Db       453  NYSRSRS 460

RESULT  11
DIG4_HUMAN STANDARD; PRT; 767 AA.
ID   DIG4_HUMAN
AC   P78352; 092941; Q9URK8;
DT   01-NOV-1997 (Rel. 35, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Presynaptic density protein 95 (PSD-95) (Discs, large homolog 4).
GN   DIG4 OR PSD95.
OS   Homo sapiens (human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC   NCBI_Taxid=9606;
OX   [1]
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=9743282; PubMed=9286702;
RA   Stathakis D.G., Hoover K.B., You Z., Bryant P.J.;
RT   "Human postsynaptic density-95 (PSD95): location of the gene (DIG4)
RT   and possible function in nonneural as well as in neural tissues.";
RL   genomics 44:71-82(1997).

[2]
RN   [2]
RP   REVISIONS.
RC   TISSUE=Mammary gland;
RX   Stathakis D.G., Hoover K.B., You Z., Bryant P.J.;
RL   Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

[3]
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20047407; PubMed=10582582;
RA   Stathakis D.G., Udar N., Sandgren O., Andreasson S., Bryant P.J.,
RT   Small K., Forstman-Semb K.;
RT   "Genomic organization of human DIG4, the gene encoding postsynaptic
RT   density 95.";
RL   J. Neurochem. 73:2250-2265(1999).

[4]
RN   [4]
RP   SEQUENCE OF 81-401 FROM N.A.
RC   TISSUE=Brain;
RX   Bremman J.E., Bredt D.S., Parkinson J.F., Manzano W.P., McClary J.A.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RA   "Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RA   -1- FUNCTION: INTERACTS WITH THE CYTOSOLIC TAIL OF NMDA RECEPTOR
RA   SUBUNIT. MAY BE INVOLVED IN SYNAPTIC JUNCTIONS PRIMARILY
RA   SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS TO BE
RA   ON THE PRESYNAPTIC SIDE (WAS ORIGINALLY THOUGHT TO BE
RA   POSTSYNAPTIC).
RA   -1- TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.
RA   -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
RA   -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
RA   -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
RA   -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
RA   This SWISS-PROT entry is copyright. It is produced through a collaboration
RA   between the Swiss Institute of Bioinformatics and the EMBL outstation
RA   the European Bioinformatics Institute. There are no restrictions on its
RA   use by non-profit institutions as long as its content is in no way
RA   modified and this statement is not removed. Usage by and for commercial
RA   entities requires a license agreement (See http://www.isb-sdb.ch/announce/
RA   or send an email to license@isb-sdb.ch).
RA   -----
CC   EMBL: U83192; AAC52113.1;
CC   EMBL: AF156495; AA056173.1;
CC   EMBL: U68138; AB07736.1;
CC   HSSP: P31016; IBE9.
CC   MIM: 602887;
CC   InterPro: IPR000619; Guanylate_kin.
CC   InterPro: IPR001478; PDZ.
CC   InterPro: IPR001452; SH3.

```

```

DR Pfam; PF00525; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
DR SH3 domain; Repeat.
KW SH3 domain; Repeat.
FT DOMAIN 108 194 PDZ 1.
FT DOMAIN 203 289 PDZ 2.
FT DOMAIN 356 436 PDZ 3.
FT DOMAIN 471 561 SH3.
FT DOMAIN 577 767 GUANYLATE_KINASE.
FT CONFLICT 46 46 E -> V (IN REF. 3).
FT CONFLICT 81 83 VIV -> EFR (IN REF. 4).
FT CONFLICT 399 401 GDO -> AGI (IN REF. 4).
SQ SEQUENCE 767 AA; 85429 MW; BE1019159E5B2D8 CRC64;

Query Match      0.6%; Score 8; DB 1; Length 767;
Best Local Similarity 100.0%; Pred. No. 18;
Matches      8;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

QY      633  DGLRLVND 640
         |||||||
Db       151  DGLRLVND 158

RESULT  12
DIG2_RAT STANDARD; PRT; 852 AA.
ID   DIG2_RAT
AC   Q63622; Q62939; P70548;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Channel associated protein of synapse-110 (Synaptic
DE   density protein PSD-93) (Discs, large homolog 2).
GN   Dig2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC   NCBI_Taxid=10116;
OX   [1]
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96310881; PubMed=8755482;
RA   Kim E., Cho K.O., Rothschild A., Sheng M.;
RT   "Heteromultimerization and NMDA receptor-clustering activity of
RT   Chapsyn-110, a member of the PSD-95 family of proteins.";
RL   Neuron 17:103-113(1996).

[2]
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96193770; PubMed=8625413;
RA   Bremman J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E.,
RA   Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F.,
RA   Froehner S.C., Bredt D.S.;
RT   "Interaction of nitric oxide synthase with the postsynaptic density
RT   protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";
RL   Cell 84:757-767(1996).

[3]
RN   [3]
RP   SEQUENCE FROM N.A.
RX   Irie M., Hata Y., Takai Y.;
RA   Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RA   -1- FUNCTION: INTERACTS WITH THE CYTOSOLIC TAIL OF NMDA RECEPTOR
RA   SUBUNIT. AS WELL AS POTASSIUM CHANNELS.
RA   -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
RA   -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
RA   -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
RA   -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
RA   -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
RA   This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: 049049; AAB53243.1; -  
DR EMBL: 050717; AAC52643.1; -  
DR EMBL: 053368; AAB48562.1; -  
DR HSSP: 012959; 1PDR.  
DR InterPro: IPR000619; Guanylate\_kin.  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00625; Guanylate\_kin; 1.  
DR Pfam: PF00595; PDZ; 3.  
DR Pfam: PF00018; SH3; 1.  
DR SMART: SM00072; Gukc; 1.  
DR SMART: SM00228; PDZ; 3.  
DR SMART: SM00326; SH3; 1.  
DR PROSITE: PS00856; GUANYLATE\_KINASE\_1; 1.  
DR PROSITE: PS50052; GUANYLATE\_KINASE\_2; 1.  
DR PROSITE: PS50106; PDZ; 3.  
DR PROSITE: PS50002; SH3; 1.  
KM SH3 domain: Repeat.  
FT DOMAIN 98 184 PDZ 1.  
FT 193 279 PDZ 2.  
FT DOMAIN 421 501 PDZ 3.  
FT 536 606 SH3.  
FT DOMAIN 662 852 GUANYLATE\_KINASE.  
FT 181 182 VR -> IL (IN REF. 2).  
FT 228 228 I -> M (IN REF. 2).  
FT 228 228 R -> K (IN REF. 2).  
FT 339 339 D -> E (IN REF. 3).  
FT 450 454 MISSING (IN REF. 2).  
FT 464 465 GD -> RK (IN REF. 2).  
FT 474 474 D -> H (IN REF. 2).  
FT 476 476 R -> P (IN REF. 2).  
FT 478 478 A -> D (IN REF. 2).  
FT 486 486 AAA -> LP (IN REF. 2).  
FT 506 506 A -> S (IN REF. 2).  
FT 569 569 H -> N (IN REF. 2).  
FT 586 586 L -> Q (IN REF. 2).  
FT 641 641 MISSING (IN REF. 2).  
FT 639 639 K -> A (IN REF. 3).  
FT 726 726 F -> L (IN REF. 1).  
FT 733 733 N -> Y (IN REF. 2).  
FT 749 749 E -> V (IN REF. 1).  
FT 756 756 L -> H (IN REF. 2).  
FT 791 792 KR -> NG (IN REF. 2).  
FT 794 794 T -> M (IN REF. 2).  
SQ SEQUENCE 852 AA; 94934 MW; F8D41AAB9CF5B09 CRC64;  
  
Query Match 0.6%; Score 8; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=66310881; PubMed=8755482;  
RA Kim E., Cho K.-O., Rothschild A., Sheng M.;  
RT "Heteromultimerization and NMDA receptor-clustering activity of  
RT Chapsyn-110, a member of the PSD-95 family of proteins.";  
RL Neuron 17:103-113(1996).  
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR  
CC SUBUNIT AS WELL AS POTASSIUM CHANNELS.  
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: U32376; AAB04949.1; -  
DR HSSP: Q12959; 1PDR.  
DR MIM: 603583; -  
DR InterPro: IPR000619; Guanylate\_kin.  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00625; Guanylate\_kin; 1.  
DR Pfam: PF00595; PDZ; 3.  
DR Pfam: PF00018; SH3; 1.  
DR SMART: SM00072; Gukc; 1.  
DR SMART: SM00228; PDZ; 3.  
DR SMART: SM00326; SH3; 1.  
DR PROSITE: PS00856; GUANYLATE\_KINASE\_1; 1.  
DR PROSITE: PS50052; GUANYLATE\_KINASE\_2; 1.  
DR PROSITE: PS50106; PDZ; 3.  
DR PROSITE: PS50002; SH3; 1.  
KM SH3 domain: Repeat.  
FT DOMAIN 98 184 PDZ 1.  
FT 193 279 PDZ 2.  
FT DOMAIN 421 501 PDZ 3.  
FT 536 606 SH3.  
FT DOMAIN 680 870 GUANYLATE\_KINASE.  
SQ SEQUENCE 870 AA; 97500 MW; 89C83BA0619F6F99 CRC64;  
  
Query Match 0.6%; Score 8; DB 1; Length 870;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 633 DGRRLRVND 640  
DB 141 DGRRLRVND 148  
RESULT 13  
DLAG2\_HUMAN STANDARD; PRT; 870 AA.  
AC 015700:  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Channel associated protein of synapse-110 (Chapsyn-110) (Discs, large  
DE homolog 2).  
GN DLAG2.

QY 633 DGRRLRVND 640  
DB 141 DGRRLRVND 148  
RESULT 14  
TYO3\_HUMAN STANDARD; PRT; 890 AA.  
AC 006418; 014953;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Tyrosine-protein kinase receptor TYRO3 precursor (EC 2.7.1.112)  
DE (Tyrosine-protein kinase RSE) (Tyrosine-protein kinase SKY) (Tyrosine-  
DE protein kinase DTK) (Protein-tyrosine kinase byk).  
GN TYRO3 OR RSE OR SKY OR DTK OR BYK.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.  
 RN NCBI\_TaxID=9606;  
 RP SEQUENCE FROM N.A.  
 RA Mark M.R., Scadden D.T., Wang Z., Gu Q., Goddard A., Godowski P.J.;  
 RT "RSE, a novel receptor-type tyrosine kinase is expressed at high levels in the brain.";  
 RL J. Biol. Chem. 269:10720-10728(1994).  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94150991; PubMed=8108112;  
 RT "Cloning of the cDNA for a novel receptor tyrosine kinase, Sky,  
 RL predominantly expressed in brain.";  
 RN Oncogene 9:699-705(1994).  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=95161080; PubMed=7857658;  
 RT "Isolation and characterization of the human DTK receptor tyrosine  
 RL kinase.";  
 RN Growth Factors 11:137-144(1994).  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=97318332; PubMed=9175267;  
 RT "Kajii Y., Nishimura Y., Kato M., Mizuguchi M., Saji M., Katsumoto T.,  
 RA Ohno K., Takashima S., Onodera K.";  
 RT "A tyrosine kinase-like molecule is localized in the nuclear membrane  
 RL of neurons: hippocampal behavior under stress.";  
 RN Biol. Cell 88:45-54(1996).  
 RP SEQUENCE OF 519-720 FROM N.A.  
 RA MEDLINE=94085793; PubMed=8262388;  
 RT "Polvi A., Armstrong E., Lai C., Lemke G., Huebner K., Spritz R.A.,  
 RA Gluda L.C., Nicholls R.D., Allitalo R.";  
 RT "The human TYRO3 gene and pseudogene are located in chromosome  
 RL 13q14-q25.";  
 RN Gene 134:289-293(1993).  
 RP FUNCTION: MAY BE INVOLVED IN CELL ADHESION PROCESSES, PARTICULARLY  
 IN THE CENTRAL NERVOUS SYSTEM.  
 CC -1 CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC -1 tyrosine phosphate.  
 CC -1 SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1 TISSUE SPECIFICITY: ABUNDANT IN THE BRAIN AND LOWER LEVELS IN  
 CC OTHER TISSUES.  
 CC -1 SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.  
 CC -1 SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1 SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; 005682; AAA19236.1;  
 DR EMBL; D17517; BAA0467.1;  
 DR EMBL; 018934; AAC50070.1;  
 DR EMBL; D50479; BAA21781.1;  
 DR EMBL; X72886; CAA51396.1;  
 DR HSSP; P11362; IRGK.  
 DR MIM; 600341;  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003598; Ig\_c2.

DR InterPro; IPR003600; Ig\_Like.  
 DR InterPro; IPR001245; Tyr\_Pkinase.  
 DR Pfam; PF00041; fn3\_2.  
 DR Pfam; PF00047; Ig\_2.  
 DR SMART; SM00069; pkinase; 1.  
 DR SMART; SM00069; fn3\_2.  
 DR SMART; SM00410; Ig\_Like; 1.  
 DR SMART; SM00408; Ig2c; 1.  
 DR PROSITE; PS00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;  
 KW Transferrase; Phosphorylation; Transmembrane; Signal; Repeat;  
 FT SIGNAL 1 40  
 FT CHAIN 1 890  
 FT DOMAIN 41 429  
 FT TRANSFEM 430 450  
 FT DOMAIN 451 890  
 FT DOMAIN 57 124  
 FT DOMAIN 153 210  
 FT DOMAIN 224 311  
 FT DOMAIN 315 411  
 FT DOMAIN 518 790  
 FT NP\_BIND 524 532  
 FT BINDING 550 550  
 FT ACT\_SITE 550 550  
 FT DISULFID 64 117  
 FT DISULFID 160 203  
 FT CARBOHYD 63 63  
 FT CARBOHYD 191 191  
 FT CARBOHYD 230 230  
 FT CARBOHYD 240 240  
 FT CARBOHYD 293 293  
 FT CARBOHYD 366 366  
 FT CARBOHYD 380 380  
 FT MOD\_RES 686 686  
 FT CONFLICT 259 36  
 FT CONFLICT 285 285  
 FT CONFLICT 293 293  
 FT CONFLICT 341 341  
 FT CONFLICT 812 812  
 FT CONFLICT 815 815  
 SQ SEQUENCE 890 AA; 96505 MW; F9EC675077C4E8F1 CRC64;  
 Query Match  
 Best Local Similarity 0.6%; Score 8; DB 1; Length 890;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 434 APASAPON 441  
 DB 321 APASAPON 328  
 RESULT 15  
 D1G1\_HUMAN  
 ID D1G1\_HUMAN STANDARD; PRT; 904 AA.  
 AC 012959; Q12958;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Presynaptic protein SAP97 (Synapse-associated protein 97) (Discs,  
 GN large homolog 1).  
 GN D1G1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.  
 RN NCBI\_TaxID=9606;  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=95024052; PubMed=7937897;

```

RA Lue R.A., Marfatia S.M., Branton D., Chishti A.H.;
RT "Cloning and characterization of hdlg; the human homologue of the
RT Drosophila discs large tumor suppressor binds to protein 4.1.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9818-9822(1994).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 460-555.
RX MEDLINE-96338231; PubMed-8757139;
RA Cabral J.H., Petosa C., Sutcliffe M.J., Raza S., Byron O., Poy F.,
RA Marfatia S.M., Chishti A.H., Liddington R.C.;
RT "Crystal structure of a PDZ domain.";
RL Nature 382:649-652(1996).
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NRDA RECEPTOR
CC SUBUNITS. ASSOCIATES WITH PROTEIN 4.1.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U13897; AAA50599.1; -.
DR EMBL: U13896; AAA50598.1; -.
DR PDB: 1PDR; 23-JUL-97.
DR MIM: 601014; -.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 3.
DR SMART: SM00018; SH3; 1.
DR SMART: SM00072; GUKC; 1.
DR SMART: SM00228; PDZ; 3.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ; 3.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain; Repeat; Alternative splicing; 3D-structure.
FT DOMAIN 224 310 PDZ 1.
FT DOMAIN 319 405 PDZ 2.
FT DOMAIN 466 546 PDZ 3.
FT DOMAIN 581 651 SH3.
FT DOMAIN 714 904 GUANYLATE_KINASE.
FT VARSPPLIC 669 680 EIPDDMGSKGLK -> OSFNDKRRKKNLFSRKFFPFYKKNKDS
FT EOETSDAQ (IN ISOFORM 2).
SQ SEQUENCE 904 AA; 100354 MW; B78798D6BB0920D4 CRC64;

```

```

Query Match 0.6%; Score 8; DB 1; Length 904;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 633 DGRLRVND 640
   |||||
DB 267 DGRLRVND 274

```

Search completed: July 24, 2002, 14:37:35  
Job time: 206 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:30:49 ; Search time 30.55 Seconds

(without alignments)  
4930.150 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 1356  
Sequence: 1 MKYVCPGRFRVVPCCGDG.....SQVARNLNLQTEKGRPEYS 1356

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq.032802:\*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	406	29.9	865	22	AA895798
2	406	29.9	865	22	AA895798
3	195	14.4	272	22	AA674163
4	195	14.4	273	22	AA117403
5	118	8.7	419	22	ABG04716
6	110	8.1	744	21	AA771410
7	80	5.9	193	22	ABG04714
8	60	4.4	60	22	AB843115
9	60	4.4	60	22	AA836946
10	60	4.4	90	21	AA771424
11	47	3.5	169	22	ABG04715

12	38	2.8	51	21	AA802888	Human secreted pro
13	31	2.3	86	21	AA771423	PDZ domain-compris
14	11	0.8	187	22	ABG04627	Novel human diagno
15	11	0.8	315	22	AB811149	Human PAR3 homolog
16	11	0.8	1464	22	AB861199	Drosophila melanog
17	8	0.6	41	21	AA853218	Protein containing
18	8	0.6	81	22	AA853384	Protein containing
19	8	0.6	106	22	AA853384	Protein containing
20	8	0.6	106	22	AA853384	Protein containing
21	8	0.6	128	21	AA809344	Arabidopsis thalia
22	8	0.6	136	21	AA809343	Human secreted pro
23	8	0.6	207	22	AA857625	Human homolog of D
24	8	0.6	207	22	AA857625	Human homolog of D
25	8	0.6	207	22	AA857625	Human homolog of D
26	8	0.6	208	22	AA855833	PDZ encoded domain
27	8	0.6	211	22	ABG10975	Novel human diagno
28	8	0.6	287	20	AA888295	Wheat geminivirus
29	8	0.6	303	20	AA774141	Human prostate tum
30	8	0.6	307	22	AA871576	Human olfactory re
31	8	0.6	307	22	AA872477	Human OR-like poly
32	8	0.6	344	22	AA855834	PDZ encoded domain
33	8	0.6	344	22	AA855834	Human post-synapti
34	8	0.6	344	22	AA855834	Human post-synapti
35	8	0.6	348	17	AA801437	Human cyclin-depe
36	8	0.6	364	22	AA875048	Human colon cancer
37	8	0.6	366	15	AA860550	Human developmenta
38	8	0.6	396	18	AA834666	Partial PSD-95 pro
39	8	0.6	414	22	AA836687	Mammalian two-hydr
40	8	0.6	450	19	AA846606	Tyrosine kinase as
41	8	0.6	450	20	AA829635	Human tyrosine kin
42	8	0.6	450	20	AA822130	Human tyrosine kin
43	8	0.6	510	22	AA848890	zinc finger protei
44	8	0.6	583	22	AB863033	Drosophila melanog
45	8	0.6	583	22	AA838948	Drosophila G-prote

#### ALIGNMENTS

RESULT	ID	AA895798	standard; Protein; 865 AA.
XX	AA895798;		
XX	26-JUN-2001	(first entry)	
DE	Human protein sequence SEQ ID NO:18775.		
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy.		
OS	Homo sapiens.		
PN	EP1074617-A2.		
PD	07-FEB-2001.		
XX	28-JUL-2000;	2000EP-0116126.	
XX	29-JUL-1999;	99JP-0248036.	
XX	27-AUG-1999;	99JP-0300253.	
PR	11-JAN-2000;	2000JP-0118776.	
PR	02-MAY-2000;	2000JP-0183767.	
PR	09-JUN-2000;	2000JP-0241899.	
XX	(HELI-) HELIX RES INST.		
XX	Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX	WPI; 2001-318749/34.		
XX	Primer sets for synthesizing polynucleotides, particularly the 5602		



Qy	474	SIIRDWTIGSSAIIYYKNNILPRAAIQQDRKLRAGDULIEVNGVDLYGKSGQEEVSLRS	5533
Db	339	SIIRDWTIGSSAIIYYKNNILPRAAIQQDRKLRAGDULIEVNGVDLYGKSGQEEVSLRS	398
Qy	534	TKMEGTVSLATFERQEDAFHPRELNAEPSSQMIPEKTRAEDEDIYLPDGRREFLTFEVPL	5933
Db	399	tkmegtvtslatferqeda fhpelnaepsqmqipketradediyltpdg rrefltfevpl	458
Qy	594	NDSSSAGLGYSVKCKRSKENHADLGIFVYKSTINGGAASKDRLRVNDQLIAVNGESILGK	6533
Db	459	sdsagsaglgysvkvckrskenhadlgifvks tlinggaaskdgrlrvndq llaingesi llyg	518
Qy	654	TNOGAMETLRSMSTEGKRRKMIDLIYAARRSKCNELKSGSPGPGLPIETLDDERR	7133
Db	519	tngametlrrsmstegkrrkmi dliyar rskcnelksgspgpel pletldderr	578
Qy	714	ISHSLVSGIEGLDPSRNAALSRING	740
Db	579	ishslvsgiegl dpsrnaal sring	605

RESULT	3	
AAAG74163		
ID	AAAG74163	standard; Protein; 272 AA.
XX		
AC	AAAG74163;	
XX		
DT	03-SEP-2001	(first entry)
XX		
DE	Human colon cancer antigen protein SEQ ID NO:4927.	
XX		
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;	
XX	colorectal carcinoma.	
OS	Homo sapiens.	
XX		
PN	WO200122920-A2.	
XX		
PD	05-APR-2001.	
XX		
PF	28-SEP-2000; 2000WO-US26524.	
XX		
PR	29-SEP-1999; 99US-0157137.	
XX	03-NOV-1999; 99US-0163280.	
XX		
PA	(HOMA-) HUMAN GENOME SCT INC.	
XX		
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;	
XX		
DR	WPI; 2001-235357/24.	
XX	N-PSDB; AAH33594.	
XX		
PT		
XX		
PS	Claim 11; Page 6691-6693; 9803pp; English.	
XX		
CC	AAH32963 to AAH37195 and AAG73514 to AAG77788 represent human colon	
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where	
CC	the proteins are collectively known as colon cancer antigens. The colon	
CC	cancer antigens have cytosolic activity and can be used in gene	
CC	therapy and vaccine production. N and P may be used in the prevention,	
CC	diagnosis and treatment of diseases associated with inappropriate P	
CC	expression. For example, N and P may be used to treat disorders	
CC	associated with decreased expression by rectifying mutations or deletions	
CC	in a patient's genome that affect the activity of P by expressing	
CC	inactive proteins or to supplement the patient's own production of P.	
CC	Additionally, N may be used to produce the colon cancer-associated Ps,	
CC	by inserting the nucleic acids into a host cell and culturing the cell	
CC	to express the proteins. N and P can be used in the prevention, diagnosis	
CC	and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204	
CC	and AAH77789 represent sequences used in the exemplification of the	

CC present invention.  
CC N.B. pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
SQ Sequence 272 AA:

Query Match	14.48;	Score 195;	DB 22;	Length 272;
Best Local Similarity	100.0%;	Pred. No. 6.8e-194;		
Matches 195;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	323	CIVRINDODLNRNRPEQAOHMFROAMRPIIWFHVPVPAANKEXEYEQULSQSKNNYYSR	38
Db	65	CIVrInddglNmrIrfegqHmfrqamrPrlWfHvvpaaNkexEqLsqsknnYysrI	124
QY	383	SPDSOYIDNRSVNSAGLFTYVORAPLNPPEOIOSHSRLPHSAPHSGKPPAPASAPONV	44
Db	125	spdegyIdnrsvnsagLftvqraprLmnppeIdshsrLlphsaPhsgkppaPasapOnv	18
QY	443	FSTVVSQGYNTKRIKRLNLOLKGTEGAGFSITSRDVTIGGSAPITYKNLLPGAAIOD	50
Db	185	fstvvsqgynTckIgrkrlnldlKkgteglgfsIstrdvtIgtsgsapIyknllpgaaId	24
QY	503	GRKAGDRLEIYNGV 517	
Db	245	grlKagdrLleivngv 259	

RESULT	4
AAU17403	
ID	AAU17403 standard; Protein; 273 AA.
XX	
AC	AAU17403;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Novel signal transduction pathway protein, Seq ID 968.
XX	
KW	Neuroprotective; cytosolic; dermatological; immunosuppressive; tumour;
KW	antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW	immune system disorder; rheumatoid arthritis; inflammatory condition;
KW	organ transplant rejection; infection; hepatitis C; blood disorder;
KW	sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW	neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW	chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW	cardiovascular; respiratory; wound healing; endocrine; Addison's disease
KW	reproductive system; gastrointestinal; liver disorder; AIDS;
XX	acquired immune deficiency syndrome.
XX	
OS	Homo sapiens.
XX	
PN	WO200154733-A1.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01312.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231442.  
PR 08-SEP-2000; 2000US-0231443.  
PR 08-SEP-2000; 2000US-0231444.  
PR 08-SEP-2000; 2000US-0231445.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234223.  
PR 25-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI: 2001-465460/50.  
N-PSDB: AAS27320.  
  
Novel polypeptides useful for diagnosing, treating, preventing and/or  
prognosing disorders related to the proteins, including cancers, immune  
disorders and neuronal disorders -  
Claim 1; SEQ ID No 968; 880bp; English.  
  
The invention relates to novel isolated polypeptides (I), and  
polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
diagnosing, preventing and treating diseases including immune system  
disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
transplant rejections and graft versus host disease, infectious diseases  
(e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
other blood-related disorders (sickle cell anaemia), myeloproliferative  
disorders, primary haematopoietic disorders, hyperproliferative  
disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal

abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g. glomerulonephritis), cardiovascular disorders (e.g. arrhythmia), respiratory disorders, dermatological disorders, in wound healing, epithelial cell proliferation, endocrine disorders (e.g. Addison's disease), reproductive system disorders, gastrointestinal disorder (inflammatory disorders), liver disorders (cirrhosis), as stimulators of B-cell responsiveness to pathogens, activators of T-cells, to induce higher affinity antibodies, and as a means to induce tumour proliferation in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction pathway protein, amino acid sequences of the invention.

Query Match 14.4%; Score 195; DB 22; Length 273;  
Best Local Similarity 100.0%; Pred. No. 6,8e-194;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 CTVRINDGDLRNRFRQAOHMFQRAHPTLIMFHVPAANKEDYEOLSOSEKNYYSSRF 382  
DB 65 CTVRINDGDLRNRFRQAOHMFQRAHPTLIMFHVPAANKEDYEOLSOSEKNYYSSRF 124  
QY 383 SPDSQYIDNRSVNSAGLHTVQARAPRLNHPEQIDSHSRPLPHSAHPSGKPPSAPAPQNV 442  
DB 125 SPDSQYIDNRSVNSAGLHTVQARAPRLNHPEQIDSHSRPLPHSAHPSGKPPSAPAPQNV 184  
QY 443 PSTTWSGQVTFKRIKRLNQLKKGTEGLCFSTSRDVTIGSAPYVKNILPRGAATD 502  
DB 185 fsltvsasygckikgrlndiqklkleglgfsltsrdvtlgsaplyvknllprgaatd 244  
QY 503 GRLKAGRLIEVNGV 517  
DB 245 grikagdrllievngv 259

RESULT 5  
ABG04716  
ID ABG04716 standard; Protein: 419 AA.

AC ABG04716;  
DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #4707.

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001MO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS68903.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 20; SEQ ID NO 35075; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 419 AA;

Query Match 8.7%; Score 118; DB 22; Length 419;  
Best Local Similarity 100.0%; Pred. No. 1.5e-113;  
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1026 FGHRKDDKTEKTKIKIOESFTSEERIRMKOERIOAKTRFRROARERDYAEIOD 1085

DB 113 fgkhrkddkletekikioesftseerirmkqeritqatrtfraqerayaeiqd 172

QY 1086 FHRTFGCDDELMYGVSSYEGSMALNARPOSPPRGHMDALYAQVKRPRNSPPSPVS 1143

DB 173 fhrtfgcddeilmgyvssyegsmalnarpqspreghmdalyaqvkkprnksppvds 230

RESULT 6  
AAV71410  
ID AAV71410 standard; Protein: 744 AA.

AC AAV71410;

DT 04-OCT-2000 (first entry)

DE PDZ domain-comprising PHIP (ephAin Interacting Protein) complex.

KW PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;  
KM cellular process; nerve cell interaction; regeneration of nerve cell;  
KW axonogenesis; antiproliferative; proliferative disorder; treatment; GRIP;  
KM differential disorder; PHIP; ephrin interacting protein; syntenin.

OS Mus sp.

PN WO200031124-A2.

PD 02-JUN-2000.

PF 19-NOV-1999; 99WO-CA01101.

PF 20-NOV-1998; 98US-0109158.

PA (MOUN ) MOUNT SINAI HOSPITAL.

PI Lin D, Pawson A;

DR WO200031124-A2.

DR 02-JUN-2000.

PF 19-NOV-1999; 99WO-CA01101.

PF 20-NOV-1998; 98US-0109158.

PA (MOUN ) MOUNT SINAI HOSPITAL.

PI Lin D, Pawson A;

Thu Jul 25 08:38:16 2002

us-09-757-781-2.olg.rag

XX WPI: 2000-400038/34.  
 DR N-PSDB: AAD01182.  
 PT Isolated complex for treating proliferative or differentiative  
 PT disorders comprises B class ephrin and PDZ domain containing protein -  
 XX Claim 26; Fig 2D; 59pp; English.  
 PS  
 CC The patent discloses a complex comprising of a B class ephrin and PDZ  
 CC domain containing protein. B class ephrins function as ligands for Eph  
 CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a  
 CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ  
 CC binding site. This complex is used in methods to modulate the interaction  
 CC of a B class ephrin and PDZ domain containing protein and to identify  
 CC modulator compounds. It is also used for modulating cellular processes  
 CC like, axonogenesis, nerve cell interactions and regeneration of nerve  
 CC cells. The complex is also useful for treating proliferative or  
 CC differentiative disorders associated with this protein complex.  
 CC The present sequence is the PHIP (ephrin interacting protein) complex.  
 CC comprising three PDZ domains. PHIP cDNA is isolated by screening the  
 CC lambdaEx10x 10.5 day mouse embryo expression library. PHIP is closely  
 CC related to Caenorhabditis elegans PAR-3 protein, involved in regulating  
 CC polarity of the early embryo. PHIP sequence can be used as a probe, to  
 CC isolate other proteins like GRIP and syntenin.

XX Sequence 744 AA;  
 SQ

Query Match 8.1%; Score 110; DB 21; Length 744;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-105; Indels 0; Gaps 0;  
 Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 39 ANDPNYWIQVHRLERHGGGILDDLLICVADDKRLVAVPEQDPHNGGSGTSSNGT 98  
 DB 39 ardpnywlvhrlengdggllldldlcvadddkrlvavpeqdpnhngsgdtsassgt 98  
 QY 99 GSPERIGSELGTNNVSAPQYQATSETEVTPSVLRAVMPHLNRRSSPAL 148  
 DB 99 gspelrsgelgtcnvsatfqyqatselevtpsvlramphlvrrssdpal 148

## RESULT 7

ABG04714  
 ID ABG04714 standard; Protein; 193 AA.  
 AC  
 XX ABG04714;  
 DT 13-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #4705.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PA (HYSE-) HYSEQ INC.  
 PI Drimanac RT, Liu C, Tang YF.  
 DR WPI: 2001-639362/73.  
 DR N-PSDB: AAS68901.  
 XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensic, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 20; SEQ ID No 35073; 103pp; English.

PS  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (II) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantifying a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 193 AA;  
 SQ

Query Match 5.9%; Score 80; DB 22; Length 193;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-74; Indels 0; Gaps 0;  
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1063 IOAKTREFRERQARERDYAEIODEFHTFCDELMYGVSSYEGSMALNRPQSPBEGH 1122  
 DB 57 lqaktrrefrerqareryaeioidfhtfcddelmlygvssyegsmalnarpspregm 116  
 QY 1123 MDALYAOYKPKRNSKPSPPVD 1142  
 DB 117 mdalyaqvkkpnskpspvd 136

## RESULT 8

ABB43115  
 ID ABB43115 standard; Peptide; 60 AA.  
 AC  
 XX ABB43115;  
 DT 04-FEB-2002 (first entry)  
 DE Peptide #10621 encoded by human foetal liver single exon probe.  
 XX  
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157277-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001; 2001WO-US00669.  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236353.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX

PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR.  
XX WPI; 2001-483447/52.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -  
XX  
XX Claim 27; SEQ ID NO 35750; 639pp + sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human fetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 60 AA;

Query Match 4.4%; Score 60; DB 22; Length 60;  
Best Local Similarity 100.0%; Pred. No. 6.4e-54;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 LVAVFEDQDPHHGDSSTGTSQSPFEGSELGTNNVSAFQPYQATSELETPSYLRA 134  
Db 1 lvaifdeqdpbhggdgtssastgtspelfgseigtvnvsafqpyqatselevpsylra 60

## RESULT 9

AAAM36946  
ID AAM36946 standard; Protein; 60 AA.  
XX  
AC AAM36946;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
DE Peptide #10983 encoded by probe for measuring placental gene expression.  
XX  
KW Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200157272-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001MO-US00663.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -  
XX  
PS Claim 27; SEQ ID NO 37215; 654pp; English.  
XX

CC The present invention relates to single exon nucleic acid probes (SENP;  
CC see AAI31315-AA157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders.  
XX  
SQ Sequence 60 AA;

Query Match 4.4%; Score 60; DB 22; Length 60;  
Best Local Similarity 100.0%; Pred. No. 6.4e-54;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 LVAVFEDQDPHHGDSSTGTSQSPFEGSELGTNNVSAFQPYQATSELETPSYLRA 134  
Db 1 lvaifdeqdpbhggdgtssastgtspelfgseigtvnvsafqpyqatselevpsylra 60

## RESULT 10

AAV71424  
ID AAV71424 standard; Protein; 90 AA.  
XX  
AC AAV71424;  
XX  
DT 04-OCT-2000 (first entry)  
XX  
DE PDZ domain-comprising PHIP PDZ3 protein complex.  
XX  
KW PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;  
KW cellular process; nerve cell interaction; regeneration of nerve cell;  
KW axonogenesis; antiproliferative; proliferative disorder; treatment; GRIP;  
KW differentiative disorder; PHIP PDZ3 complex; ephrin interacting protein.  
XX  
OS Mus sp.  
XX  
PN WO200031124-A2.  
XX  
PD 02-JUN-2000.  
XX  
PF 19-NOV-1999; 99WO-CA01101.  
XX  
PR 20-NOV-1998; 98US-0109158.  
XX  
PA (MOUN ) MOUNT SINAI HOSPITAL.  
XX  
PI Lin D, Pawson A.  
XX  
DR WPI; 2000-400038/34.  
XX  
PT Isolated complex for treating proliferative or differentiative  
PT disorders comprises B class ephrin and PDZ domain containing protein -  
XX  
XX Claim 3; Fig 2C; 59pp; English.  
XX  
XX The patent discloses a complex comprising of a B class ephrin and PDZ  
CC domain containing protein. B class ephrins function as ligands for Eph  
CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a  
CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ  
CC binding site. This complex is used in methods to modulate the interaction  
CC of a B class ephrin and PDZ domain containing protein and to identify  
CC modulator compounds. It is also used for modulating cellular processes  
CC like, axonogenesis, nerve cell interactions and regeneration of nerve  
CC cells. The complex is also useful for treating proliferative or  
CC differentiative disorders associated with this protein complex.  
CC The present sequence is the PDZ domain-comprising PHIP (Ephrin  
CC Interacting Protein) PDZ3 protein complex. PHIP consists of three PDZ  
CC domains and is closely related to Caenorhabditis elegans PAR-3 protein,  
CC involved in regulating polarity of the early embryo. This sequence  
CC is isolated by screening the lambdaExlox 10.5 day mouse embryo  
CC expression library.  
XX  
SQ Sequence 90 AA;

Query Match  
 Best Local Similarity 4.4%; Score 60; DB 21; Length 90;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 594 NDSGAGVSVKRNKSENHADIGIVKSIINGGAASKDGRIRVNDOLIAVNGESLIGK 653  
 Db 6 ndsagaglvsvkvgrnksenhadigivkslinggaaskdgrirvndqilavngesllgk 65

RESULT 11  
 ABG04715  
 ID ABG04715 standard; Protein; 169 AA.  
 AC ABG04715;  
 DT 13-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #4706.  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 PA WO200175067-A2.  
 PN 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Dmanac RT, Liu C, Tang YF;  
 PI WPI: 2001-639362/73.  
 DR N-PSDB; AAS68902.  
 XX

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 20; SEQ ID No 35074; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pcl\_sequences.

Sequence 169 AA;

Query Match  
 Best Local Similarity 3.5%; Score 47; DB 22; Length 169;  
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1144 NRSTPSNHDRIQRIROFEOAKODEVEDRRTYRSPQPPNARPAT 1190  
 Db 85 nrstpsnhdriqrlrgefqqakqdevdrtrrtystqppnarpap 131

RESULT 12  
 AAG02888  
 ID AAG02888 standard; Protein; 51 AA.  
 AC AAG02888;  
 DT 06-OCT-2000 (first entry)  
 DE Human secreted protein, SEQ ID NO: 6969.  
 KW Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KM gene therapy; chromosome mapping.  
 OS Homo sapiens.  
 PA EP1033401-A2.  
 PN 06-SEP-2000.  
 PD 21-FEB-2000; 2000EP-0200610.  
 PF 26-FEB-1999; 99US-0122487.  
 PR (GEST) GENSET.  
 PA Dumas Milne Edwards J, Duclert A, Giordano J;  
 PI WPI: 2000-500381/45.  
 DR N-PSDB; AAC02894.  
 XX

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -

Claim 13; SEQ ID 6969; 71pp + CD-ROM; English.

The present sequence is a polypeptide encoded by one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

Sequence 51 AA;

Query Match  
 Best Local Similarity 2.8%; Score 38; DB 21; Length 51;  
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1023 MFRPKHKKDKIKETGKTIQESFTSEERIRMKQD 1060  
 Db 14 mfrgkhrkdkkiketgkikiqesftseerirmkqg 51

RESULT 13



ID	AAV71423	standard; Protein: 86 AA.
XX	AAV71423	
AC	AAV71423;	
XX		
DT	04-OCT-2000	(first entry)
XX		
DE	PDZ domain-comprising PHIP PDZ2 protein complex.	
XX		
KW	PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;	
RK	cellular process; nerve cell interaction; regeneration of nerve cell;	
KM	axogenesis; antiproliferative; proliferative disorder; treatment;	
KX	differentiative disorder; PHIP PDZ2 complex; ephrin interacting protein.	
XX		
OS	Mus sp.	
XX		
PN	MO200031124-A2.	
XX		
PD	02-JUN-2000.	
XX		
PF	19-NOV-1999; 99WO-CA01101.	
XX		
PR	20-NOV-1998; 98US-0109158.	
XX		
PA	(MOUN ) MOUNT SINAI HOSPITAL.	
XX		
PI	Ltn D, Pawson A;	
XX		
DR	WPI; 2000-400038/34.	
XX		
PT	Isolated complex for treating proliferative or differentiative disorders comprises B class ephrin and PDZ domain containing protein -	
XX		
PS	Claim 3; Fig 2C; 59pp; English.	
XX		
CC	The patent discloses a complex comprising of a B class ephrin and PDZ domain containing protein. B class ephrins function as ligands for Eph receptor tyrosine kinases (RTK) and possess a transmembrane element and a highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ binding site. This complex is used in methods to modulate the interaction of a B class ephrin and PDZ domain containing protein and to identify modulator compounds. It is also used for modulating cellular processes like, axonogenesis, nerve cell interactions and regeneration of nerve cells. The complex is also useful for treating proliferative or differentiative disorders associated with this protein complex.	
CC	The present sequence is the PDZ domain-comprising PHIP (Ephrin interacting protein) PDZ2 protein complex. PHIP consists of three PDZ domains and is closely related to Caenorhabditis elegans PAR-3 protein.	
CC	Involved in regulating polarity of the early embryo. This sequence is isolated by screening the lambdaExlox 10.5 day mouse embryo expression library.	
XX		
SQ	Sequence 86 AA:	
OY	Query Match 2.3%; Score 31; DB 21; Length 86; Best Local Similarity 100.0%; Pred. No. 1.4e-23; Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DB	489 YVKNIIPGAAIIOGRLKAGDRIIVNGVDL 519                                       29 yvknllprgaalqdgrrlkagdrllveavgvd1 59	
RESULT 14		
ABG04627	ABG04627 standard; Protein: 187 AA.	
AC	ABG04627;	
XX		
DT	13-FEB-2002 (first entry)	
XX		
DE	Novel human diagnostic protein #4618.	

KW	Human: Chromosome mapping; gene mapping; gene therapy; forensic;
KV	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
PR	31-MAR-2000; 2000US-0540217.
ER	23-AUG-2000; 2000US-0649167.
PA	(HYSE-) HYSEQ INC.
PI	Dermanac RT, Liu C, Tang YT;
DR	WPI: 2001-639362/73.
XX	N-PSDB; AAS68814.
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
PS	Claim 20; SEQ ID NO 34986; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent need not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pcl_sequences.
XX	
SQ	Sequence 187 AA;
XX	
OY	Query Match 0.8%; Score 11; DB 22; Length 187;
Db	Best Local Similarity 100.0%; Pred. No. 0.019;
	Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
OY	106 SELGTNNYSAP 116
Db	36 selgtnnysaf 46
RESULT 15	
ID	ABB11149
ID	ABB11149 standard; peptide: 315 AA.
AC	ABB11149;
DT	11-JAN-2002 (first entry)
DE	Human PAR3 homologue, SEQ ID NO:1519.
XX	

Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 chronic inflammatory condition; proliferative retinopathy;  
 atherosclerosis; coronary heart disease; arterial ischemia;  
 bone disorder; osteoporosis; vascular growth disorder;  
 tissue regeneration; wound healing; infection; immune disorder;  
 cell culture; drug screening; gene therapy; antiinflammatory;  
 antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 cyostatic; osteopathic; vasotropic; cardiac; virucide; antibacterial;  
 antifungal; vulnerary; antitumor.

Homo sapiens.

WO200157188-A2.

09-AUG-2001.

05-FEB-2001; 2001WO-US03800.

03-FEB-2000; 2000US-0496914.

27-APR-2000; 2000US-0560875.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-457740/49.

N-PSDB; ABA08393.

Human proteins and DNA encoding sequences useful for preventing,  
 treating or ameliorating a medical condition in a mammalian subject  
 e.g. arthritis and cancer.

Claim 20; Page 150; 1963pp; English.

Sequences ABA08225-ABA09574 represent 1350 novel human polypeptides, and  
 sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 invention also relates to vectors and recombinant host cells comprising a  
 nucleotide of the invention, methods of producing the novel polypeptides,  
 antibodies against the polypeptides, methods of detecting the nucleotides  
 or polypeptides in a sample, and methods of identifying compounds which  
 bind to polypeptides of the invention. Although novel, many of the  
 polypeptides of the invention have homology to known proteins, thereby  
 giving an insight into their probable biological activities, and hence  
 potential therapeutic applications. The polypeptides of the invention may  
 have various activities, including cytokine, cell proliferation or cell  
 differentiation activities; stem cell growth factor activity;  
 haematopoietic regulatory activity; tissue growth activity;  
 immunomodulatory activity; activin- or inhibin-related activities;  
 chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 thrombolytic activities; receptor or ligand activities; or may be  
 involved in oncogenesis, cancer cell proliferation or metastasis.  
 Depending on their biological activities, polypeptides and nucleotides of  
 the invention are useful for preventing, treating or ameliorating medical  
 conditions, e.g., by protein or gene therapy. Such conditions include  
 cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 proliferative retinopathy, atherosclerosis, coronary heart disease,  
 arterial ischemia, bone disorders (e.g., osteoporosis), and abnormal  
 vascular growth. Polypeptides involved with tissue regeneration and  
 repair (or nucleic acids encoding them) may be used to promote wound  
 healing (e.g., of burns, incisions and ulcers), while those with  
 immunomodulatory activities may be used in the treatment of viral,  
 bacterial and fungal infections in addition to immune disorders.  
 Polypeptides with growth factor activity may be used in cell cultures to  
 promote cell growth. For example, such polypeptides may be used to  
 manipulate stem cells in culture to give rise to neuroepithelial cells  
 that can be used to augment or replace cells damaged by illness,  
 autoimmune disease or accidental damage. The polypeptides and nucleotides  
 may also be used in the diagnosis of the above conditions, and in drug

CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.

XX Sequence 315 AA;

Query Match 0.8%; Score 11; DB 22; Length 315;

Best Local Similarity 100.0%; Pred. No. 0.03;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 GGPLGIHVVF 291

DB 39 GGPLGIHVVF 49

Search completed: July 24, 2002, 14:34:44  
 Job time: 235 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 06:39:23 ; Search time 3972.89 Seconds  
(without alignments)  
19327.033 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689

Sequence: 1 atgaagtgacggtgctt.....tcttcctgttaaaaaaaaa

Scoring table: IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlun:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	936.6	16.5	1078	9	AL529631
C 2	886.2	15.6	1074	9	AL529864
C 3	765	13.4	870	10	BG743714
C 4	761.2	13.4	767	10	BG1917860
C 5	757.4	13.3	1150	10	BM466811
C 6	734.6	12.9	1063	10	BM462581
C 7	714.6	12.6	852	10	BG745599
C 8	710.8	12.5	725	10	BE792557
C 9	686.6	12.1	713	10	BG170486
C 10	665.8	11.7	1040	10	BG745723
C 11	665	11.7	973	9	AL529632
C 12	658.2	11.6	694	10	BG910244
C 13	644.6	11.3	760	10	BG195361
C 14	644	11.3	945	9	AL529865
C 15	638.2	11.2	966	10	BG260547
C 16	637.8	11.2	941	10	BG336977
C 17	626.4	11.0	687	9	BE176633

C 18	605.2	10.6	645	10	BE693704
C 19	603.4	10.6	779	10	BG209378
C 20	600.8	10.6	985	10	BG122324
C 21	599.2	10.5	987	10	BG170054
C 22	599	10.5	608	9	AW452651
C 23	594	10.4	617	10	BG431070
C 24	586.4	10.3	599	9	AW780417
C 25	579.8	10.2	618	9	BE176440
C 26	573.2	10.1	592	9	AL599694
C 27	573	10.1	660	10	BG473458
C 28	571.4	10.0	574	10	BG470230
C 29	557.8	9.8	814	9	AU079784
C 30	554.6	9.7	561	9	AU079784
C 31	553.4	9.7	718	10	BG206726
C 32	553	9.7	575	10	BG1870563
C 33	552.6	9.7	626	10	BE679962
C 34	549.6	9.7	740	10	BG184956
C 35	538.4	9.5	699	9	BS455982
C 36	536.8	9.4	971	10	BG261114
C 37	536	9.4	536	9	AU079538
C 38	536	9.4	536	9	AU079538
C 39	536	9.4	809	10	BI093826
C 40	530.4	9.3	540	9	AW242451
C 41	526.2	9.2	545	9	AI129988
C 42	515	9.1	535	10	BE870124
C 43	511.4	9.0	534	9	AW835321
C 44	507.4	8.9	600	10	BE299145
C 45	504.6	8.9	1207	10	BF581891

#### ALIGNMENTS

RESULT 1  
AL529631 1078 bp mRNA linear EST 13-FEB-2001  
LOCUS AL529631 LTI\_NFL001\_NBC4 Homo sapiens cDNA clone CS0DD0051D18 3  
DEFINITION prime, mRNA sequence.  
ACCESSION AL529631  
VERSION AL529631.1 GI:12793124  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1078)  
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL  
TITLE  
AUTHORS  
COMMENT  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

#### FEATURES

source  
1..1078  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DD0051D18"  
/clone\_1lb="LTI\_NFL001\_NBC4"  
/sex="male"  
/tissue-type="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@lifestech.com URL: http://fulllength.invitrogen.com" 6 others  
BASE COUNT 236 a 283 c 222 g 331 t

## ORIGIN

Query Match 16.5%; Score 936.6; DB 9; Length 1078;  
Best Local Similarity 97.7%; Pred. No. 5.5e-154;  
Matches 982; Conservative 3; Mismatches 7; Indels 13; Gaps 3;

QY 2081 ggaacccccctggagacgtgagcccaatgaaacagcgttgatgataagaacgaaga 2140  
|||||  
DB 1067 GGAGCCCCCTGGACCTG-GTGGCCATGAKACAGCTTGATGATGAGGAAGAA 1009  
|||||  
QY 2141 ttcccatctccctacagtgaggttgagggttgatgatacgcacgaagaatctg 2200  
|||||  
DB 1008 TTTCCATTCCCTCTACAGTGGGATGAGGGCTTGATGATGCCCGCAGGAATGCTG 949  
|||||  
QY 2201 ccttaagagataatggtgagtcaggtlaaatacacagcttccctacagtgatagc 2260  
|||||  
DB 948 CCGTCAGTAGATATG-----GGTAAATACCRCTGTCTCCCTACAGTGAATGTC 898  
|||||  
QY 2261 cccaagatgacacgtgcattatagaagaatgacaggttgccagtgctccctccacatctc 2320  
|||||  
DB 897 CCCAAGATGACACTGCTCATATGAGAGATGACAGGTTGCCAGTCTCTCCACATCTCT 838  
|||||  
QY 2321 ctgacacgtcccttccacagctcccatgatatgtgtgggtttgtgacggcagatgctgta 2380  
|||||  
DB 837 CTGACAGTCTCTCTCCSGCTCCCATGATGATGAGGTTGTGTGACGGCAGATGCTGTA 778  
|||||  
QY 2381 ctgaggccaaggtcgaatcgaatgtatcagccgactgctcttgatgcagatgtgac 2440  
|||||  
DB 777 CTGGGGCAGGCTGCATACAGATGATTCAGCGACTGCTCTTGATGCCAATGTTGATC 718  
|||||  
QY 2441 caagtcttgcttcttcaacagaaagtgtgacgctcagatgatacagaataaacacaa 2500  
|||||  
DB 717 CAGTTCTGCTTTTCAACGAGAAAGATTGACGTGAGATGTCAGAAAAACCCACAA 658  
|||||  
QY 2501 agcaatttccagatgcagtcgaatgtgattgtgtttaaacaacagaatcaaaagacatg 2560  
|||||  
DB 657 AGCAATTTTCAGATGCCAGTCATGATGATTTCTTAAACACGAAATCAAAAAGCATGG 598  
|||||  
QY 2561 attaggtatagctgcagagagactaaactcaatacagtgatgacccgaagaagcaggtctc 2620  
|||||  
DB 597 ATTATG---TTGCTGACGAGACTAACTCAATACAGTGAACCGAAGACAGGTTCTC 541  
|||||  
QY 2621 ccaagcagatgtgtgtcttccctccttggtgtgaaagaatcgaatcgtgttgagagtcgc 2680  
|||||  
DB 540 CCMGCAAGATGTGGTCTCTTCCCTGGGTCTGAAGAAAGTCAAGCTCTTGAGAGTCTGC 481  
|||||  
QY 2681 agacgcagttgcagaggttgacttgaaatgggatatctcttccatcgtccacggcgc 2740  
|||||  
DB 480 AAGACGAGTTGCCGAGGTGACTTTGAATGGGATATTCCTTCCATGCTCCACGGCGC 421  
|||||  
QY 2741 ggataatcagagcagggatgacatgagagcttcagagctgcacatgcacaatcttatg 2800  
|||||  
DB 420 GGATTAATCGAGGCAAGGAGATCAATGAGRGTTTCAAGGGCTGCCATGCATCAAACTTATG 361  
|||||  
QY 2801 ataaacccgcgtatagatgatatgaaagcactgagacacttggaaaagaacacagaag 2860  
|||||  
DB 360 ATTAACCCGGGTAGATGATGATGAAGCGATGAGACCTTGGAAGAAAGACACAGAG 301  
|||||  
QY 2861 aaagtccaagatcagagagagagctgtatccacagcagtgatcagcttcccaatctc 2920  
|||||  
DB 300 AAAGTTCAATCAGAGAGAGAGTCTGTATCCACAGCCAGCTGATCAGCTTCCACTCTC 241  
|||||  
QY 2921 tggagacacaaatgaaatgaaacaaagagaaagtgtataagactgtatgaaaaaagagta 2980  
|||||  
DB 240 TGGAGAGCAAAATGAAATGGAACCAAGAGAAAGTGATTAAGACTGATTAAGAAAAAGATTA 181  
|||||  
QY 2981 aaactgaaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 3040  
|||||  
DB 180 AAACCTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 121  
|||||  
QY 3041 aggaatcgtgaaggtcttggaagacatgtcaggtttgcaaac 3085  
|||||

Db 120 AGGAATGCTGAAGGCTTGGAGACATGTTCAAGCTTGCCAAAC 76

## RESULT 2

AL529864/c 1074 bp mRNA linear EST 13-FEB-2001  
LOCUS AL529864.L1.NFL001.NBC4 Homo sapiens cDNA clone CSDD005YH02 3  
DEFINITION prime, mRNA sequence.  
ACCESSION AL529864  
VERSION AL529864.1 GI:12793357  
KEYWORDS EST.  
SOURCE human.  
ORANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1074)  
L.I., W.B., Gruber, C., Jesse, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.

## FEATURES

## source

1. 1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CSDD005YH02"  
/clone\_1id="L1.NFL001.NBC4"  
/sex="male"  
/tissue.type="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
cloned, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
vector. Library was normalized. Library was constructed  
by Life Technologies. Contact: Feng Liang Life  
Technologies, a division of Invitrogen 9800 Medical Center  
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
8371 Email: filiang@lifetech.com URL :  
http://fulllength.invitrogen.com"

BASE COUNT 229 a 289 c 221 g 324 t 11 others

## ORIGIN

Query Match 15.6%; Score 886.2; DB 9; Length 1074;  
Best Local Similarity 96.5%; Pred. No. 3.6e-145;  
Matches 932; Conservative 8; Mismatches 13; Indels 13; Gaps 3;

QY 2120 tggatgataagaaagaaatccctacagtgatgaggtgagctgag 2179  
|||||  
DB 1011 TGGATGATAGAGAACG-AGAAATTCGAATTCCTTCAAGTGGGATTTGAGGGCTTGATG 953  
|||||  
QY 2180 aatgcacagcgaatgctgcctcagtaggataatggtgagtcaggttaattaccagc 2239  
|||||  
DB 952 AATGCCCGCAGGAATGCTGCTCAGTAGATAATG-----GTTAAATACCMGC 902  
|||||  
QY 2240 tgtccctcagtgatagatgcccacagatgacatgcatattatagaagatgacaggttc 2299  
|||||  
DB 901 TGTCCCTCTCAATGATATGATGCCCAAGATGACACACTGATATAGAAGATGACAGGTTGC 842  
|||||  
QY 2300 cagtgcttcctcacatctctctgaccagctcctctccagctcccatgatatgtgggt 2359  
|||||  
DB 841 CAGTGCTTCTCCACATCTCTGACCAATGCTCTTCVCGCTCCATGATGATGAGGCT 782  
|||||  
QY 2360 ttgtacgacagatgctgtgacttggtggccaagctgcgaatcagtgatcacaagcagctgct 2419  
|||||  
DB 781 TTGAGACGCGAGATGCTGTGACTTGGCCAAAGCTGCAATTCAGTATCAGCCGACTCT 722  
|||||  
QY 2420 ctgtgagtcagatgtgtatccagttctgtcttcaacgagaagatgttgagctcaga 2479  
|||||  
DB 721 CTTGATCCAGATGTTGATCCAGTTCTTGCTTTCACAGAGAGATTGAGACGTCA 662  
|||||

QY	2480	gtatgctcgaataaacgcgcgaagaacattttcaagatgcgaagtcatttgatttcglttaaa	2539
Db	661	gtttcttcgaaanmngccacaaagcaatttttcgattgcacgattcaattggattttgaaaa	602
QY	2540	cacgaaatcctaaaacatctgaattgaagtlatgctgcagacgactaaactcaatacagttg	2599
Db	601	CACGAAATCAGAAAAGCATGGATTWGG---TTGCTGACGAGACTTAACCTCATACAGTGG	545
QY	2600	atgaccagaagaacaggtctccacagcagaagatgtyggtcttcacctggtctgaagaagt	2659
Db	544	ATGACCAAGAAAGCAGAGTTCTCCGCGCAAGATGTGGTCTTCCTCGGCTCGAAGAAAT	485
QY	2660	caagctgtgtgagagctctgcagaccgaagtgtccgaagtgtaacttgaatggtgatatcc	2719
Db	484	CAAGCTCGTTGGAGAGTGTGCAGACCGCAGTGTCGAGGTGACTTTGGAATGGGATATTTC	425
QY	2720	ctttccatcgtcccaagcgcgcgataatcagaagcagggatgtaagtgaagcttcagag	2779
Db	424	CTTTCCATCGTCCACGGCCGCGAATATCAGACGAGGGGATGCATAGAGGCTTCAGGG	365
QY	2780	ctgcacatgcacaatccttatgataaaaccgcggtatgatatgatatgaagcattgaga	2839
Db	364	CTGCGATCGAACAAMCTTWTGATTAACCCGCGGTAGATGATGATGATGAAGCATAGAGA	305
QY	2840	ccttggaaagaagacacagaagaagaatttaagaatcagggagagagatctatccaaagcca	2899
Db	304	CTTGGAAAGGAGACACAGAAAGATTGAAAGTCAGGAGAGAGTCTATCCACAGCCA	245
QY	2900	gtgatcaagccttccacactctctgagagacaaatgaatgaaaccagaagaagttgata	2959
Db	244	GTGATCACCTTCCACCTCTCTGGAGACACAATTAATGGAACCAAGAAAGGTGATA	185
QY	2960	agactgatagaaaaaagataaaactgtgaaaaaagaagaagaatagatagaataaggaga	3019
Db	184	AGACTGATAGAAAAAGGATTAAGCTGAAAAAGAAAAAGAAATGAGATTAAGAGAGA	125
QY	3020	aggatlaaataaaaagccagaagaaggaatgctggaagggcttggaaacatgttcaagtttg	3079
Db	124	AGGATAAATATAAAGCCAGAAAGGAATGCTGAAGGGCTTGGAGACATGATCCACCTTGG	65
QY	3080	gcaaac 3085	
Db	64	CCAAAC 59	
RESULT	3		
LOCUS	BG743714	870 bp	EST 15-MAY-2001
DEFINITION	602632895F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4778302 5',		
ACCESSION	BG743714		
VERSION	BG743714.1	GI:14054367	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.		
WORDS	1 (bases 1 to 870)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: James Cleaver, M.D.		
	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA		
	Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LHAM10633 row: m column: 23		
	High quality sequence start: 5		

High quality sequence stop: 844.									
FEATURES		Location/Qualifiers							
SOURCE		1. 870							
		/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4778302" /clone_11b="NCI CGAP_Skn3" /lab_host="DH10B (T1 phage-resistant)" /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI_CGAP Library."							
BASE COUNT		206 a 208 c 213 g 243 t							
ORIGIN									
	Query Match	13.4%;		Score 765;		DB 10;		Length 870;	
	Best Local Similarity	97.5%;		Pred. No. 6e-124;					
	Matches 841;	Conservative 0;		Mismatches 15;		Indels 7;		Gaps 6;	
OY	4311	tcatactctctgagcctctccacagtagtaacaggaatctggccctactctatagatgttcg	4370						
Db	4	TCATATGCTCTGGCATCTCCACAGTACCGAATCTCGG-CCTATCTATGAGTGGCGCG	62						
OY	4371	cggcttggcttgtaaatccctctgacacacttgcagtagaacaattcacctgaagtgaagat	4430						
Db	63	CGGCTTGGTGTAAATCATCCCTGCACACATCTGCAGTGCAGCAAAATTCACCTGAATGGAGGAT	122						
OY	4431	gaactgagccctgtttctctcctctaagtctctcttagctatgagtagacatcttagctc	4490						
Db	123	GACGTGGCGCCCTGGTTCCTCCTTAAGTTCCTTAAGCTATGGATGACATCTTAATCTTC	182						
OY	4491	tgtgtgaggaagaagtgggcgacatacaccaaaattggggcttctgtgacttccacaga	4550						
Db	183	TGTGTGAGAGAAAGTGGCGGACATACACCAAAATTTGGGCTTTCTGTACTTCCACAGCA	242						
OY	4551	cagccattgtctgaacttgtcatcactgtgtgttctctcttctcttctcaagctcttgt	4610						
Db	243	CAGCCATTTGTGCTACTTTGTCATCAGTCTGCTTTTCCTTTCCCTTTCACACTCTTTGT	302						
OY	4611	gaacggagaagtgcgtcatctcctattcaagaagttaagcaatgccaactgtttgtgtca	4670						
Db	303	GACGGGAGAGTGGTCATCTCTATTACAGAGCTTAAGCCATATCTCAACATTTGTGGTCA	362						
OY	4671	coatgggggtcccttctgttaactgccttaagctacaacttacaataaagtatgatcct	4730						
Db	363	CCATGGGGGTCTCTTTGTAACTGCGTTAAGTCAACATTTACCAATTAAGTGATCTCT	422						
OY	4731	ggtctgggttataataatagctgttgttcgttctctgttctctgaaagtgggttaagtcaca	4790						
Db	423	GGTCTGGCTTTAATACATACGCTTGTTCGGTCTCTGTTCTCGAACAGTGGGTTAAGTCACCA	482						
OY	4791	cagctctgtgtgggaaacgttggagacagaaatgtgctcctcgcgaggggagagcttggcctg	4850						
Db	483	CAGCTCTGTGTGGGAAGCTGGGAACACAGAGATGGCTCTCTGCCGGGGGAAGCTGGGCTCG	542						
OY	4851	ccattggccctgtgtctatcatagaggaggagcctaagaaagaattctcttagaagagc	4910						
Db	543	CCATTGGCCCTGTGTCTATCATAGAGGGGAGACCTAAGAAAGAAATTTCTCTAGGAAGAC	602						
OY	4911	tcatbgcccaatcaactcttagtaattatttaattagtttggctctgacacattgttcg	4970						
Db	603	TCATBGCCCACTACCTCTAGTAATTAATTTTAATTAAG-TTTTGTCTTGACACTTGTCTAG	661						
OY	4971	gaaggagcaagaatggagc--agagataaacacagacagtcatttgaatctgtctctaa	5028						
Db	662	GAAAGGACACAGATGGGACACAGATTAAACACAGACAGCATTTTGATGTGCTCTAGC	721						
OY	5029	gttttcaagtcagaaggcaattgagcttgttcttaatgcatccaacactgcatgtctgac	5088						
Db	722	GTTTTTCAGTCTAGAGGC-ATTGATGCTTGTCTTAATGCAATCCACACACTGCAGTCTGAC	780						
OY	5089	tggcgatgcacacgctccctaaagtatgtctgcacatgaacaacataaagaacaaaggaaaagccg	5148						

Db 781 TGG-GATGCCACGCTCTTAAGTACTTCTGCCATGCAACAT-AAAGACAAGGAAAGCGG 838  
 QY 5149 ttacacatcacacagagaacatt 5171  
 Db 839 TTACACATCACACAGACACATTT 861

RESULT 4  
 BI917860 767 bp mRNA linear EST 16-OCT-2001  
 LOCUS 60318390F1 NIH\_MGC\_121 Homo sapiens cDNA clone IMAGE:5247718 5',  
 DEFINITION mRNA sequence.  
 ACCSSION BI917860  
 VERSION BI917860.1 GI:16199788  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 767)  
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 TITLE NIH-MGC http://mgi.nci.nih.gov/  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

FEATURES  
 High quality sequence stop: 759.  
 Location/Qualifiers

1..767

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5247718"  
 /clone\_1id="NIH\_MGC\_121"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pCMV-SPORT6; Site: 1; NotI;  
 fetal brain; female age 20 weeks; female age 24 weeks;  
 and male age 26 weeks. Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.7 kb, insert size range  
 0.7-3.5 kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research Genetics tracking code 017. Note:  
 this is a NIH\_MGC Library."  
 BASE COUNT 236 a 179 c 190 g 162 t  
 ORIGIN

Query Match 13.4%; Score 761.2; DB 10; Length 767;  
 Best Local Similarity 99.6%; Pred. No. 2.8e-123;  
 Matches 763; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1115 gtgagaagaacattaccattcaagccgttttagccctgacagcagattatgtgcaaaa 1174  
 Db 1 GTGAGAAGAACATTACTATTCAAGCCGTTTAGCCCTGCACGACGATATTATGACAAACA 60  
 QY 1175 ggaagtgttaaaagtgcagggtctacacagctgcagagagacccccgactgaaccacccgc 1234  
 Db 61 GGAGGTGAACAGTGCAGGGCTTCACACGGTGTGAGAGACACCCGACTGAAACACCCGC 120  
 QY 1235 ctgagcagatagactctcaactcaagactccctacagcagacacccctcggaagaaacac 1234  
 Db 121 CTGAGCAGATAGACTCTCACTCAAGACTACTGCTATAGGCAACCCCTCGGGAACACAC 180  
 QY 1295 catccgctcagcagctcgacacctcgaatgtattagtagcagctgtgaagctgtgtata 1354

Db 181 CATCCGCTCAGCCCTCGCACCTCAGATGATTTAGTACGACTGTAAACAGTGTATA 240  
 QY 1355 acaccaaaaaataaggcaagaggttataatccagcttaagaagggtacagaagttgg 1414  
 Db 241 ACACCAAAAAATAGGCAAGAGCTTTATATTCACGCTTAAGAAAGTACAGAGGTTGG 300  
 QY 1415 gattcagatcactctcagagatgtatacaatagtggtcctcagctcaatctatgaana 1474  
 Db 301 GATTCACTACTCCTCCAGAGATGAACAATAGGTGGCTCACTCAATCTATGTGAAA 360  
 QY 1475 acattctcccccggggggggccatcagagtgccgacttaagcagagacagactta 1534  
 Db 361 ACATTCTCCCCCGGGGGGGCCATTTCAGATGGCCGCACTTAAGCAGAGACGACTTA 420  
 QY 1535 taaggttaattggaatgtaagttaagggaataatccagaagaagatgtgttcgttga 1594  
 Db 421 TAGAGTAAATGAGTACATTTAGTGGGCAAAATCCAAAGGAAGTGTGTTGCTGTGA 480  
 QY 1595 gaagcaccagaatggaaggaactgtgagcctctgtgtcttcgcagagaagacgttcc 1654  
 Db 481 GAAGCAACCAAGTGGAAAGAACTGTGAGCCTCTGTGCTTCCGAGGAAGACGCTTCC 540  
 QY 1655 acccaagggaactgaatgcagagccagccagatgcagattccaaaagaagcaagcag 1714  
 Db 541 ACCCAAGGAAGTCAATGTCAGAGCCAGCCAGATGCAGATTCGCAAAAGAAAGCAAGCAG 600  
 QY 1715 aagttgagatattgtctctacacctgagtgccagaggaatttgcattgaagtc 1774  
 Db 601 AAGATGAGGATATGCTCTTCACTGATGGCAGCAAGGAATTTCTGCAATTTGAATGC 660  
 QY 1775 caattatgattcagatctgcagcgtgtgtcagtgtaagaagcaagcagcag 1834  
 Db 661 CACTTATGATTCAGATCTGACAGCCTTGGTGTAGTCAAGTCAACGGTCAACAAAG 720  
 QY 1835 agaaccagcagattggaatcttctgaagtcattatcaagtg 1880  
 Db 721 AGAACCAACCAATCTGGGAATCTTTGCAAGTCAATTTACTAGC 766

RESULT 5  
 BM466811 1150 bp mRNA linear EST 05-FEB-2002  
 LOCUS AGNCOURT\_6431318 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:5503072  
 DEFINITION 5', mRNA sequence.  
 ACCSSION BM466811  
 VERSION BM466811.1 GI:18515853  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1150)  
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 TITLE NIH-MGC http://mgi.nci.nih.gov/  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

High quality sequence start: 9  
 High quality sequence stop: 652.  
 Location/Qualifiers

1..1150  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5503072"

```

/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORE6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT      294 a      310 c      309 g      235 t      2 others
ORIGIN

```

```

Query Match      13.3%; Score 757.4; DB 10; Length 1150;
Best Local Similarity 97.0%; Pred. No. 1.2e-122;
Matches 804; Conservative 0; Mismatches 21; Indels 4; Gaps 3;

OY 3333 cgcctagacccctagagcccccagagggagatagatgtagcttctgtatgcccagatcaa 3392
|||||
DB 38  CGCTGAGACTCTGAGACCCACGAGAAAGGCGATGTGATGATGCTTTGTATGCTCAAGTCAA 97
|||||

OY 3393 gaagccgcggaaattccaaacccctcacctctagacagtaacatcactctagaacatca 3452
|||||
DB 98  GAAGCCGGCGGAATTCCTCAAAACCTCACCTGTAGACAGTAAACATGATCAACTCTAGCAATCA 157
|||||

OY 3453 tgatcgatagagcgctcgtgaggaagaatttcagcaagaagaagaatgtagatgtaga 3512
|||||
DB 158  TGATCGGATACGCGCTGTGAGGCAAGAAATTTTCAGAACGCAAGATGAAGATGTAGA 217
|||||

OY 3513 agatcgctgcggagacctaagttttgagcaaccctgcgcgaacgcagcgcgcgcgcgcga 3572
|||||
DB 218  AGATGCTGGCGGCGGACCTATAGTTTGTAGACAAACCTGCGGAAACGACGCGCGGACGCA 277
|||||

OY 3573 gaagcggcgcaactcgtcgtcgtcgtagaggtgcagatgcagcgcgcgcgcgcgcgcgcgc 3632
|||||
DB 278  GAGCGCGGCGACACTGCTGCTCGGTGAGGTGCAGATGACAGCGCGACGCGCAGAGAGAGCG 337
|||||

OY 3633 cgaagagctcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3692
|||||
DB 338  CGAGAGGCTCCAGCAGAGGCCAGCGCAGACGACTCTGCTCGGCAAGCAAGCAAGCAAGCA 397
|||||

OY 3693 tgcacagctcgtctccacaggaactctctggagagagaactcaccctctggggaagcttcca 3752
|||||
DB 398  TGCCAGGCTCGGTCTCCAGGACTCTTGGAGACGAGAACTACTCCCTGGGGAAGGCTTCCA 457
|||||

OY 3753 gactgcgaagaagaccccaaggtactcagatcaccaggtctccaggaacgcgttaccctgg 3812
|||||
DB 458  GAGTCCCAAGAGAACCCACAGGTACTCCAGCTAACAGGCTCCAGAGACGCGTACTCTGGG 517
|||||

OY 3813 aggacatggtctcaacgcgcgaaggtcactgtgaaactcaagagctccttcgcgcgaaga 3872
|||||
DB 518  AGGACATGAGCTTCAACGCGCAGGGTCACTGTAAGAACTCAGAGACTCTTCCGCAAGACAA 577
|||||

OY 3873 gaagcggagagagcagcagatgaaagaacagcctcctccgcggggcccaagaactatga 3932
|||||
DB 578  GAGGCGGAAGAGCAGCAGATGAAGAAGAGCGCTCTCCGAGGGGCCAGAACATATGA 637
|||||

OY 3933 ctctgataagaagaatccagagcccaagtttaagccctcccaagggg -cccttcggcgaag 3991
|||||
DB 638  CTCGTATGAAGAAATTCAGGAGCCAGGACCCCTCCCTCCCGGCGGCCCTTCCGCAAG 697
|||||

OY 3992 atgtt -ccccccctccctctcaggttcgaaagctgaaacaagaactcagaactctaga 4050
|||||
DB 698  ATGTGCCCCCTCCCTCTTCTCAAGTTGCGAGGCTGMAACAGACTTCAGACTCTTGAAAA 757
|||||

OY 4051 -ggagagccctctatctctcctgagcagcaaaataacagatgtctatgtcgcacataa 4108
|||||
DB 758  AGGAGAGCCCTTCTTCTTCTGAGCATGAATAACCGATGCTTATATGCCCCGCAATAA 817
|||||

OY 4109 agacatttctcatgaagaactgtatcttgggaagttttttaaacctc 4157
|||||
DB 818  AGACATTTTCTATGAAGACTTGTATTTTGGAGTTTTTTAAACC 866
|||||

```

RESULT 6

```

BM462581      1063 bp      mRNA      linear      EST 05-FEB-2002
LOCUS
DEFINITION
AGENCOURT_6426369 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5518290
5', mRNA sequence.
ACCESSION
BM462581.1 GI:18511621
VERSION
EST.
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1063)
NIH-MGC http://mgi.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
UNPUBLISHED (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cga@h-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM12178 row: n column: 19
High quality sequence stop: 565.
Location/Qualifiers
1. 1063
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5518290"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomysarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pCMV-SPORE6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT      307 a      247 c      259 g      250 t
ORIGIN

Query Match      12.9%; Score 734.6; DB 10; Length 1063;
Best Local Similarity 93.9%; Pred. No. 1.2e-118;
Matches 855; Conservative 0; Mismatches 39; Indels 17; Gaps 8;

OY 1723 gatattgtcttaaccatgtagcaccaggaattctcagattgaatccactaat 1782
|||||
DB 1  GATATTGTTCTTACCTGATGCGACCAAGGAATTTCTGACATTTAAGTCCACTTAAT 60
|||||

OY 1783 gattcagatctgcagagccttggtcagtgctcaaaagtaaccggtcaaaagagaaaccac 1842
|||||
DB 61  GATTGAGATGTCGACAGCGCTTGCTGTCAGTGTCAAGAGTAACCGGTCAAAAGAGAACAC 120
|||||

OY 1843 gaagatttggagatcttctcaagtcattatataagagggagcagacatctaagatga 1902
|||||
DB 121  GCGAGATTGGGAATCTTGTTCAGAGTCCATTATTAATGAGAGCACATCTTAAGATGA 180
|||||

OY 1903 aggcctcggtgaatgatcaactgatagcagtaaatggaagaaacccctgttggcgaagaa 1962
|||||
DB 181  AGGCTTCGGGTGAATATCACTGATAGCAATGAAGAGAAATCCCTGTTGGCAAGACA 240
|||||

OY 1963 aaccaagatgcattgaaaccccaagaaggtctatgtctactgaaggaataaagaagaa 2022
|||||
DB 241  AACCAATATAGCCATGGAACCCCTAAGAAAGTGTATGTCTCAAGCAATTAACAGAGA 300
|||||

OY 2023 atgattcaactatgttttcaagaggaataagaagaatgaatgaatgaatgaatgaatga 2082
|||||
DB 301  ATGATCCAGCTTATTTGTTCAGAGGAATTAAGAAATGAAGTGAAGTCACTGAGG 360
|||||

OY 2083 agcccccctgcagccttagctgcacatgaaacagcgtctgtagatagaagaagaatc 2142
|||||
DB 361  AGCCCCCTTGAGACTGAGGTGCCCATTTGAACAGCCTTGGATGATGAGAACAGAAATTT 420
|||||

```

```

FEATURES
source
1 of 3
Location/Qualifiers
stop: 829.
sequence column: 0

```

note="Organ: spleen; Vector: pOTB7; site\_1: XhoI; site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned"

	Query Match	12.6%	Score 714.6;	DB 10;	Length 852;
	Best Local Similarity	92.5%;	Pred. No. 3,9e-115;		
	Matches 839;	Conservative	0;	Mismatches 9;	Indels 59; Gaps
Dy	1946 cccgttggcagaacaaagatgcaaatgcatcttggaaccttaagaagtgcattcttcacg				
Dd	2 CCGTGTGGGCAAGACAAACCAGAATGCCGTGAACCCCTTAAGAAAGGTCTATGTCTACAGC				
Oy	2006 aaggcaataaacagagagatatccagcttatgtgtgcagaagagataaagaagctgaatg				
Dd	62 AAGGCATAAATACAGGAATGATCCAGCTTATTGTTCMAAGAGAMATACCAAGTCAATG				
Oy	2066 agctgaagttactcctggagaccoccccttgagcttagctgcocatlttaaacagcgltgatg				
Dd	122 AGCTGAAGTCCCTGGAGCCCCCTTGACCTTAGCTGCCCTTTGAAACACCGTTGGATG				
Oy	2126 atgaaaacagaaatlctccattccctctacaagttggattggaagggcttgatgaatgc				
Dd	182 ATAGAGAACGAAGAAATTCCTCATCTCCCTCMACAGTGGGATTTGAGAGGCTTGATGAATGC				
Oy	2186 ccagagaanaatgctgcctcctagtaggataatgygtgcagttaataacacagcgtgcc				
Dd	242 CCAAGAAAATGCGCCCTCAGTAgAATAG-----GATAAATACAGCTGTGCC				
Oy	2246 ctacagtaaatatgcccacaaagatgacactgcatltaaagaatagacagcttgccaatgc				
Dd	293 CTMACAGTAATATGCCCCAMAGATGACACTGTCTATTATGAACATCACAGGTGCCATGCC				
Oy	2306 ttcctcacatctctcgaacagtlccctcttcagctccatccatgatatggttgttga				
Dd	353 TTCCCTCCACATCTCTGACACAGTCTCTTCCAGCTCCCATGATGATGGGGTTGTGA				
Oy	2366 cggcagatgctgacttggccaagctgcgaatcagttgataacagcgtcgtcttga				
Dd	413 CGGCAAGTCTGTACTTGGCCAMGCTGTGCAATCATGATTCACGCACTGCTCTTTGA				
Oy	2426 gtccagatgttalcacagtlcttgcatttccaagaagaagatttggacgtcagatgt				
Dd	473 GTCCAAATGTATTCAGTCTTCTTCTTTCACAGAAAGATTTGGACGTCAAGATNTGT				
Oy	2486 cagaaaaacgcaaaacaaatlctcagatgcagatcattgatttgcgtlaaaacagaa				
Dd	533 CAGAAAAACGCAAAACCATTTTTCAGATGCCAGTCAATTTGGATTTCGTTAAACAAGAA				
Oy	2546 aatcaaaaacagatgatttagtlatagctgcagactaaactcaatcacagygatgacc				
Dd	593 AATCAAAAACATGATTTAGCTA-----				
Oy	2606 agaagcaggttctccagcagagatgttggtccctccctggcttgaagaagtcaagct				
Dd	617 -----GTTCTCCACGACAGATGTGGGTCTTCCCTGGGCTCAAAACATCAAGCT				
Oy	2666 cglttgagaatctcgaacagccagattg -cegaagtgcatttgaatggatattcc-ttt				
Dd	668 CATGGAAAGTGTGCAGAACCGCAGTGTGCCCAAGGTGACTTTGAATGGGATATTCCTTTT				
Oy	2724 caatctgcacagccgcgagataatcagaagcagagagtgataatgagagcttoagaagtcg				
Dd	728 CCATCTGCTCCACGGCCCGGATCTCTAAGAGCAGGGGATCCATATGAGACTTCAAGAGCTGC				
Oy	2784 catcgacaatcttattgataaccgcgylagattgatatgataagacalvgagacctt				
Dd	788 CATGCACAAATCTTATGATTAACC--GCCGTATGATGATGATGAAGCAT--GAGACTTT				



Oy 2844 ggaagaa 2850  
 Db 845 GGAAGAA 851  
 RESULT 8  
 BE792557 725 bp mRNA linear EST 20-SEP-2000  
 LOCUS 601585341P1 NIH\_MGC\_7 Homo sapiens CDNA clone IMAGE:3939370 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE792557  
 VERSION BE792557.1 GI:10213755  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 725)  
 AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: DCTD/DRP  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: L10M788 row: b column: 11  
 High quality sequence stop: 725.  
 Location/Qualifiers  
 1..725  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3939370"  
 /clone\_1lb="NIH\_MGC\_7"  
 /tissue\_type="small cell carcinoma"  
 /cell\_line="MGC3"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; CDNA made by oligo-dt priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-CDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 256 a 118 c 207 g 144 t  
 ORIGIN  
 Query Match 12.5% Score 710.8; DB 10; Length 725;  
 Best Local Similarity 99.6% Pred. No. 1.8e-114;  
 Matches 723; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
 Oy 2597 tggatgaccagaagcaggttcccccagcagagatgtgtgtctccctcgtggtcgaaga 2656  
 Db 1 TGGATGACCAGAAACAGGTCTCCACGACAGAGATGTGGTCTTCCTCGGCTCGAAGA 60  
 Oy 2657 agtcaagctcgttggagagcttgcagaccgagctgtccgagggtgaattgaatggagga 2716  
 Db 61 AGTCAAGCTCATTTGGAGAGTGTGACAGACCGCAGTTGCCAGAGTGACTTTGAATGGGATA 120  
 Oy 2717 ttcccttcacatcgtccaaagcgccggagataatcagaagggagaggaatgaatgaatc 2776  
 Db 121 TTCTTTCATCTGCTCCAGCGCGGATATCATGAGCGAGGATGCAATGAGACTTCA 180  
 Oy 2777 gagctgcacatcgacaactcttataaaccgcggtatgatgatgatgaaggcatg 2836  
 Db 181 GAGCTGCATGACAAATCTTATGTATTAACCCGCGGTAGATATATATGTAAGCATG 240

Oy 2837 agaccttgaagaagaacacagaagaagatccaagatccaggagagagctglatccacag 2896  
 Db 241 AGACCTTGAAGAAGACACAGAAAGTTCAAGATCAGAGAGAGCTGTATCCACAG 300  
 Oy 2897 ccaagtatcagcttcccaactctcttggagaagacaattgaatggaaaccagaagagatg 2956  
 Db 301 CCAGTGATCAGCTTCCCACTCTCTGGAGACAAATGATGAAACCAAGAGAAAGGTG 360  
 Oy 2957 atcaagctgatagaagaagagataaactcgtgaagaagaagaagaatagatgaag 3016  
 Db 361 ATTAAGACTATGAAAAAAGATTAACCTGAAAAAGAAAGAAATAGATAGAG 420  
 Oy 3017 agaagataaattgaagccaaaggaagatcgtgaaggcttggagacatgttcaggt 3076  
 Db 421 AGAAGATTAATAATGAAGCCAAAGCAAGATGCTGAAGGCTTGAGACATGTCAGGT 480  
 Oy 3077 ttggcaaacatcggaaaagttgacaagattgagaagaacgggttaataaataacaggaat 3136  
 Db 481 TTGGCAAAACATGMAAAAGTGAACAGATTGAGMAACGGGTTAAATAATACAGGAAT 540  
 Oy 3137 ccttaccatcagaagagagagatagatgaatgaagcagagcagagagattcaagcca 3196  
 Db 541 CCTTACATCAGACAGAGAGAGATACGAATGACAGACAGACAGAGATTCAAGCCA 600  
 Oy 3197 aaactcgaatttaggaagcagcaagctcgaagcgttgactatgctgaattcaagatt 3256  
 Db 601 AAACTCGAATTTAGGAAAGCAAGCTCGAGAGCGTCACTGTGAAATCAAGATT 660  
 Oy 3257 ttcatctgacatttgctgtgatgatgaattaatgtatggggagcttcttataagaag 3316  
 Db 661 TTCATCGACATTTGG-TGTGATGATGAGATTATGTATGGGGAGTCTTCTTATGAAG 719  
 Oy 3317 gtccca 3322  
 Db 720 GTTCCA 725  
 RESULT 9  
 BG170486 713 bp mRNA linear EST 06-FEB-2001  
 LOCUS BG170486  
 DEFINITION 602322859P1 NIH\_MGC\_89 Homo sapiens CDNA clone IMAGE:4425975 5',  
 mRNA sequence.  
 ACCESSION BG170486  
 VERSION BG170486.1 GI:12677189  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 713)  
 AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM10172 row: e column: 16  
 High quality sequence stop: 710.  
 Location/Qualifiers  
 1..713  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4425975"  
 /clone\_1lb="NIH\_MGC\_89"  
 /tissue\_type="hypermephroma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: kidney; Vector: pCMV-SPORT6; Site\_1: NotI;

us-09-757-781-20.rst

Page 8

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 1040)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Genome  
Center  
Unpublished (1999)

FEATURES  
 Cloned and sequenced by: Incyte Genomics, Inc.  
 Found through the I.M.A.G.E. Consortium (LNL)  
<http://image.llnl.gov>  
 Plate: LICM1691 row: 0 column: 07  
 High quality sequence start: 15  
 High quality sequence end: 15

Location/Qualifiers  
1. .1040

BASE COUNT		NH <sub>2</sub> -MGC Library. <sup>a</sup>	(White Technologies).	Note: This is a
ORIGIN	231 a	265 c	220 g	324 t

Accession	Indels	Gaps
1017	10	7
1018	10	7
1019	10	7
1020	10	7
1021	10	7
1022	10	7
1023	10	7
1024	10	7
1025	10	7
1026	10	7
1027	10	7
1028	10	7
1029	10	7
1030	10	7
1031	10	7
1032	10	7
1033	10	7
1034	10	7
1035	10	7
1036	10	7
1037	10	7
1038	10	7
1039	10	7
1040	10	7
1041	10	7
1042	10	7
1043	10	7
1044	10	7
1045	10	7
1046	10	7
1047	10	7
1048	10	7
1049	10	7
1050	10	7
1051	10	7
1052	10	7
1053	10	7
1054	10	7
1055	10	7
1056	10	7
1057	10	7
1058	10	7
1059	10	7
1060	10	7
1061	10	7
1062	10	7
1063	10	7
1064	10	7
1065	10	7
1066	10	7
1067	10	7
1068	10	7
1069	10	7
1070	10	7
1071	10	7
1072	10	7
1073	10	7
1074	10	7
1075	10	7
1076	10	7
1077	10	7
1078	10	7
1079	10	7
1080	10	7
1081	10	7
1082	10	7
1083	10	7
1084	10	7
1085	10	7
1086	10	7
1087	10	7
1088	10	7
1089	10	7
1090	10	7
1091	10	7
1092	10	7
1093	10	7
1094	10	7
1095	10	7
1096	10	7
1097	10	7
1098	10	7
1099	10	7
1100	10	7
1101	10	7
1102	10	7
1103	10	7
1104	10	7
1105	10	7
1106	10	7
1107	10	7
1108	10	7
1109	10	7
1110	10	7
1111	10	7
1112	10	7
1113	10	7
1114	10	7
1115	10	7
1116	10	7
1117	10	7
1118	10	7
1119	10	7
1120	10	7
1121	10	7
1122	10	7
1123	10	7
1124	10	7
1125	10	7
1126	10	7
1127	10	7
1128	10	7
1129	10	7
1130	10	7
1131	10	7
1132	10	7
1133	10	7
1134	10	7
1135	10	7
1136	10	7
1137	10	7
1138	10	7
1139	10	7
1140	10	7
1141	10	7
1142	10	7
1143	10</	

2259 gccccaagatg--aaactat  
2259 gccccaagatg--aaactat

2315 atctctctg-accagctcctcttccagctcccatcatgatatctccac 838  
|||||  
837 attctctctt |||||  
attctctctt |||||

2434 gttatccacgctggcgcagcgttcgaatttgcgcacgat 2433

2493 acgcacaagcaat-ttcagatgccaattacaa  
|||||TTCACGAGAGAGATTGGACGTCAAGAATATGCAGAAAA 658

...AACGTTAAACACGMAAATCTA 598  
-aaagcattgattagtatatgcctgacgagactaacatcatcacgtatgatccca  
597 AAGCATGATGATTAAAA

.....ccggaagtcagctcggttg 2671

Db	579	--GTTCC	CCGACGAGAGATGTGGGTCTTCCCTCCCTGGGCTCAAGAAATCAAGTCATTTGG	523
Qy	2672	agaatctcgaacccgcaagtctccgaagtgaacttggaaatggagataatcttcacatgctc	2731	
Db	522	AGAGTCTGCAACACCCGCAATCTCCGAGTGAATGATGATGAAGCATGAGACCTTGGAAAG	463	
Qy	2732	cacgacccgagataatcgaagcgaaggatgacatgaaagcttcgaagcttcacatcgaca	2791	
Db	462	CACGGCCCCGGATATATCATCAGAGGCAAGGATTCATGAGACCTTCAGACCTGCCATCGACA	403	
Qy	2792	aactctatgataaaccgcggttagatgataatgaaagcatggaaccccttggaaag	2851	
Db	402	AATCTTATGATTAACCCCGGTAGATGATGATGATGAAGCATGAGACCTTGGAAAG	343	
Qy	2852	acacagaagaagaatcgaatcgaagcgaagcgtctgtatccacagccagatgaagctt	2911	
Db	342	ACACAGAAGAAAGTTCAAGATCAGGAGAGAGTGTATTCACACCCGATATCAGCCTT	283	
Qy	2912	cccaactctctggagagacaataatgtaattgtaaaaccaagaagaagtgataagctgaagaa	2971	
Db	282	CCCACCTCTCTGAGACGACAAATGAATGCAAAACCAAGCAAAAGGATGATAGCTGATAGAA	223	
Qy	2972	aaaaggataaaacttggaaaagaagaagaagaatagataaaggaaaggaataaataatga	3031	
Db	222	AAAAGGATTAACACCTGGAAGAAAGAAAGAAAGATAGAGATAGAGAGAAAGATATAAATGA	163	
Qy	3032	aagccaagaagaatgctgtaagcgtctggagacatgttcaagcttggcaaacatcgaa	3091	
Db	162	AAGCCAAAGAGGAATCCTGAAGGCTTGGAGACATGTTCAGCCTTCCCAAAACTGGAAG	103	
Qy	3092	aagatgacagaattgagaagaacggtaaaataaataat	3128	
Db	102	CCCGAGAGAGATGACAAACAAACGATTCAAACAT	66	
RESULT 11				
AL529632		973 bp	mRNA	linear
LOCUS	AL529632	LTI_NFL001_NBC4	Homo sapiens	CD0005YD18 5
DEFINITION	AL529632	prime, mRNA sequence.		
ACCESSION	AL529632			
VERSION	AL529632.1	GI:12793125		
KEYWORDS		EST.		
SOURCE		human.		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
TITLE		L1.W.B., Gruber, C., Jessee, J. and Polayes, D.		
JOURNAL		Full-length cDNA libraries and normalization		
COMMENT		Unpublished (2001)		
FEATURES		Contact: Genoscope		
source		Genoscope - Centre National de Sequencage		
		Bp 191 91006 EVRY cedex - France		
		Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.		
		Location/Qualifiers		
		1..973		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/clone="CS0DD005YD18"		
		/clone_id="LTI_NFL001_NBC4"		
		/sex="male"		
		/tissue_type="neuroblastoma cells"		
		/lab_host="DH10B"		
		/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA		
		was primed with a NotI-oligo(dT) primer. Five prime end		
		cloned, double-stranded cDNA was digested with Not I and		
		cloned into the Not I and Eco RV sites of the pCMVSPORT 6		
		vector. Library was normalized. Library was constructed		
		by life technologies. Contact : Feng Liang life		
		Technologies, a division of Invitrogen 9800 Medical Center		
		Drive Rockville, Maryland 20850, USA Fax : (1) 301 610		
		8371 Email : fliang@lifestech.com URL :		

BASE COUNT	209 a	290 c	286 g	171 t	17 others
ORIGIN					
Query Match	11.7%; Score 665; DB 9; Length 973;				
Best Local Similarity	97.3%; Pred. No. 1.9e-106;				
Matches	674; Conservative 11; Mismatches 7; Indels 1; Gaps 1;				
QY	1	atgaagtgaccgtgcttcggacgcgacccgggtgctgctgacctgctgcyggygaacgccac	60		
Db	281	ATGAAGTGAAGTCGCTKCTTGATGGACGCACCGGGGTGCTGCTGCCGTGGCGGACGGCCAC	340		
QY	61	atgaagtcttcacgctcatcgcgaaggcggtgaccgcctacgcgaagaagccatgcgaag	120		
Db	341	ATGAAGAKTTTTCAGCCTCATCCAGCAGCGGGTAGACCCTCACCGGAAGCCATTCCCAGA	400		
QY	121	gatccaactcttgtatcacgltgcatcgcttggacaactltagaatgaggaatactaaag	180		
Db	401	GATCCAACCTACTGGATTCAGGTGCTATGCTTTGGAACTGAGATGAGAACTACTAAGAC	460		
QY	181	cttgatgacattctttgtatgtagcaagataaagaacaagacttgtagcagtgcttat	240		
Db	461	CTTGATGACATTCTTTGTGATGTACACAGCGATAAAMKACAGCTGTTGACAGTTTTGAT	520		
QY	241	gagcagatccacatcacggaagtgatgaccagtgccagtlcccaggtccagggtaeccagagc	300		
Db	521	GAGCAGGATTCACAATCACGAGGATKATGGCAACCACTKCAGTTCCACGGGTACCCAGAGC	580		
QY	301	ccagagatatttggtagtgagcttggaccaaacatgtctcaagcctttcagccttaccaa	360		
Db	581	CCAGAGATATTTGGTAGTTAGCTTGGCACCAACAATGTCACGCTTTCAGCCTTAACAA	640		
QY	361	gcaacaagtgaatatgtagagtcacacccttcagtccttcgacgaacaatgctcttcatt	420		
Db	641	GCAACAATGTAATGAGAGTGCACACCTTCAGTCCCTTGACGAATAATGCTCTTCATGTT	700		
QY	421	cgaagcagtagtgaccagctctaattgagcctctccactctgtcagtgatagtaattt	480		
Db	701	CGAGCACAATGTTGACCCAGCTCTAATTTGGCCCTCCTCACCTTGTCTAGATAGTDATTTT	760		
QY	481	tcccttgagaagagccttcaagaataaacacacgccttgtaaacaagccttggctcttc	540		
Db	761	TCCCTGTAAGAGCCTTCAAGGAAAAATCCCAACCTGCTCAACAACAGCTGGCTTCTC	820		
QY	541	aagcagaacactgct-gggagtcctcaaaaacctgycacaggaagaagaatgaanaactacag	599		
Db	821	AAGCAGAACACTGCTGGGGAGTCCCTMAAACCTGCGACACAGAAAGATGAAAATCTACAG	880		
QY	600	aagcctcccgcggaatactagtaactggtctlaaccaatlccagagagaacaatgctgctc	659		
Db	881	AAGCCTCCCCGGGAGTACTAGTAAGTGGCTPMAACCAATTTGAGAGAGACAATGCTGCGTC	940		
QY	660	gtctctgagtgccagtcacccaatgvtggcaaa	692		
Db	941	GTCCTGTGAGTGCACAGTCAACCACATGTGTGGGAR	973		
RESULT 12					
LOCUS	BG910244	694 bp	mRNA	linear	EST 05-JUN-2001
DEFINITION	60280573.F1 NCI_CGAP_Brn67 Homo sapiens cdNA clone IMAGE:4937865				
ACCESSION	BG910244				
VERSION	BG910244.1 GI:14290720				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Mammalia: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia: Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 694) NIH-MGC htcp://mgs.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC)				

## JOURNAL COMMENT

Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: David N. Louis, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium (LNL).  
 http://image.llnl.gov  
 Plate: LLM10872 row: f column: 10  
 High quality sequence stop: 694.  
 Location/Qualifiers  
 1. 694

## FEATURES

source  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="4937865"  
 /clone\_lib="NCI\_CGAP\_Brn67"  
 /tissue\_type="neoplastic oligodendroglioma with 1p/19q loss"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: brain; Vector: PCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."  
 BASE COUNT  
 186 a 189 c 205 g 114 t

## Query Match

Best Local Similarity 11.6%; Score 658.2; DB 10; Length 694;  
 Matches 685; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

OY 3280 gataagtaatgatatgggagagttcttcttataagaagttccatggtccacgctaga 3339  
 DB 1 GATAGTTAATGTATGGGAGTTCTTCTATGAAGTGCCAGCTCTCAACGCTTAA 60  
 OY 3340 cctcagaagcccaagaagggcagatgatgatgtcttctatcccaagttcaagagcg 3399  
 DB 61 CCTCAGAGCCACGAGAGAGGCGATATGATGATCTTTGATCCCAATCAAGAGCCG 120  
 OY 3400 cgaatcccaaacctcactcctgtagacagtaaacagatcaactcctagcatatgatcg 3459  
 DB 121 CGGATTTCCAAACCTCCTACCTGTAGACATGACATCACTCTTACCATCATGATCG 180  
 OY 3460 atacagcgtctgaagcaagaatttcagcaagaagcaagatgataagatcgt 3519  
 DB 101 ATACAGGCTCGAGGCAAGATTTCAGCAAGCAAGCAAGATGAAGATGAGATCGT 240  
 OY 3520 cggcgaacctatagtttgaagcaacctggccgaagcagcgagcgagcagcgagcg 3579  
 DB 241 CGGCGAGACCTATAGTTTATGCAACCTCGCCGAAGCGACGCGCGAGAGCGG 300  
 OY 3580 cgacactcggttcggtgaggtgacagatgacgagcagcagcagcagcagcagc 3639  
 DB 301 CGGCACTCGGTGCTGCGTGGAGCTGCAAGTGCAGCGCGCAGCAGAGCGCGAGAC 360  
 OY 3640 tcccaagcagcccaagcagcagatagctctgctcctcggcaagcaggaagaaatgcaagc 3699  
 DB 361 TCCCGACAGGCGCCAGCCAGTACGCTCTGCTCCGCAAAAGAGAGAAAATGCCAGC 420  
 OY 3700 tgggtctcccgagctcttggaagcagaactacccctggggaagaggttccagagtgc 3759  
 DB 421 TCGGCTCCCGAGGACTTTGGAGAGCAAC7ACTCCCTGGGGAAAGGCTTCCAGAGTGC 480  
 OY 3760 aaagaagaaccccggttaactcagctaccagagctcccggaagcagctacggtggaagac 3819  
 DB 481 AAAGAGAACCCCGAGTACCTCCAGTACCAAGCTCCAGGAACCGCTACCTGGAGAGACAT 540  
 OY 3820 ggttcaagcagaggtcagctgct-ggaactcaggaagctcctcgcgcaggaagcagagcg 3878  
 DB 541 GGCTTCAAGCGAGGCTATCTGGAAGTCAAGAGCTCTTCCGCAAGAGACAGAGCGG 600

## RESULT 13

LOCUS BGI95361 760 bp mRNA linear EST 21-APR-2001  
 DEFINITION RST14543 Athersys RAGE Library Homo sapiens CDNA, mRNA sequence.  
 ACCESSION BGI95361  
 VERSION BGI95361.1 GI:13717048  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

## REFERENCE

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Harrington J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R., E., Veloso, N., Kika, A., Hess, J., Cottrien, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M.  
 Creation of genome-wide protein expression libraries using random activation of gene expression  
 Nat. Biotechnol. 19 (5), 440-445 (2001)  
 MEDLINE 21227151  
 COMMENT Contact: Scott J. Cain  
 Athersys, Inc.  
 3201 Carnegie Ave, Cleveland, OH 44115, USA  
 Tel: 216 431 9900  
 Fax: 216 361 9596  
 Email: scain@atersys.com

## FEATURES

source  
 Location/Qualifiers  
 1. 760  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Athersys RAGE Library"  
 /note="line="H1080"  
 /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression' Nature Biotechnology, in press. Note that even though the cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances."  
 BASE COUNT  
 258 a 139 c 197 g 164 t 2 others

## Query Match

Best Local Similarity 11.3%; Score 644.6; DB 10; Length 760;  
 Matches 653; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 2419 tcttgagtcagatgattgatacattctcttcaagcagaagattgagctcag 2478  
 DB 7 TCTTACAGTCCAGGTGTTTCCAGTTCTTCTTTCACGAAAGATTTGAGCTAG 66  
 OY 2479 agtatgcagaaacacacaaagcaatttcagatgacagtcagtcagtcagtcag 2538  
 DB 67 AGTATGTCAGAAAAACGCAAAACGATTTTCAGATGCCAGTTCGATTGCTTAA 126  
 OY 2539 acacgaataatcaaaaacatgagattagatagctgagcagacactaaactaatacagtg 2598  
 DB 127 ACACGAATAATCAAAAACATGATTTAGCTATAGCTGACGAGACTAAATCAATACAGTG 186  
 OY 2599 gatgacagaagaagcagttctccagcagaagatggtggtctcctccctgaggtctgaaagag 2658  
 DB 187 GATGACTAGAAACAGGTTCTCCAGAGACATGCTGCTCTTCCCTGGGTATCAAGAG 246

2659 tcaagctcgttgagagatctcagacgcagattccagagtgacttgatcagtgagataatt 2718  
|||||  
Db 247 TCAAGCTCGTGTGAGAGCTCAGACCGCAGTTCGAGTGACTTTGCAATGGGATATT 306  
2719 ccttccatcgtccagcgcgcgataatcagagcagggatggaatgagagcttca 2778  
|||||  
Db 307 CCTTCCGTCCGCCCGCCGCCGATATCAGAGCGAGGATGCAATGACAGCTTCAGA 366  
2779 gctgcacatcgaacaattatgatataaaccgcggttagatgatgatgaagcagc 2838  
|||||  
Db 367 GCTGCATCGCAAAATCTTATGATTAACCCCGGTAGATGATGATGAAGGATGAG 426  
2839 accttgaagaagacacagaaagttcaagatcagagagagagctcgtatccagcc 2898  
|||||  
Db 427 ACCTTGAAGAAGACACAGAAAGATTCAAGATCAGGAAGAGATCTGTATCCACACC 486  
2899 agtatcagccttcccatctcgtgagagacaatgatgaaaccaaagagaagatgat 2958  
|||||  
Db 487 AGTATTCAGCCTTCCCATCTCTGTGAGACAAATGAAATGGAACCAAGAGAAAGTAT 546  
2959 aagactgatagaataaagataaacttgaataaagaaagaaagaaagataagag 3018  
|||||  
Db 547 AAGACTGATAGAAAAGATTAACCTGAAAAGAAAAGAAAGATAGATAGAGAG 606  
3019 aagataaataagaaagcagaaagaggaatgctgaaggcttggagacatgttcagttt 3078  
|||||  
Db 607 AAGGTAATAAGTAAAGCCAGAAAGGGAATGCTGAAGGCTTGGAGACAAATTCAGCCTT 666  
3079 ggcaaac 3085  
|||||  
Db 667 GCCAAG 673

RESULT 14  
AL529865 945 bp mRNA linear EST 13-FEB-2001  
LOCUS AL529865 LTI\_NFL001\_NBC4 Homo sapiens cDNA clone CSDD005YH02 5  
DEFINITION prime, mRNA sequence.  
ACCESSION AL529865  
VERSION AL529865.1 GI:12793358  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 945)  
AUTHORS Li, W. B., Gruber, C., Jessup, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
SOURCE  
Location/Qualifiers  
1..945  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CSDD005YH02"  
/clone\_lib="LTI\_NFL001\_NBC4"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-Oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@life.com URL : http://fulllength.invitrogen.com"

BASE COUNT 204 a 284 c 286 g 164 t 7 others

ORIGIN  
Query Match 11.3%; Score 644; DB 9; Length 945;  
Best Local Similarity 98.5%; Pred. No. 8.8e-103;  
Matches 656; Conservative 4; Mismatches 5; Indels 1; Gaps 1;  
1 atgaagtgaccgtgtgtcttcgagcgaacccgggtgtgtcgtgtcgggagcgcac 60  
Db 280 ATGAAGTGACCAGT 339  
61 atgaagtttccagcctatccagcagcggtgtgtaccggtctacgggaagggccatc 120  
Db 340 ATGAAGTTTCCAGCCTCATCCAGCAGCGGAGCCCGTACCGGAAGCCATCCCAAG 399  
121 gatccaactacttgatatacaggtgatcgtcttgagacatgagatgaggaatactag 180  
Db 400 GATCCAAACTACTGATATACAGGTGCATCGCTTGGAAATGAGATGAGGAATFATGAC 459  
181 ctgtatgacatcttctgtatgtatgtagcagacgaataaagacagactgtgtatgt 240  
Db 460 CTGTATGACATCTTGT 519  
241 gagcagatccacatcaacgaggtgtatgtgacccaggtgccagttccagggtaacca 300  
Db 520 GAGCAGATCCACATCAGCAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 579  
301 ccagagataattgtgtatgtatgtgtgacccaacaaatgtctcagccttcagcctta 360  
Db 580 CCAGAGATATTGT 639  
361 gcaacaagtgaataatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtat 420  
Db 640 GCAACAAGTGAATAATGT 699  
421 cgaagcagtagtgagaccagcttaattgtgctcctcactcctcgtatgtatgtatgt 480  
Db 700 CGAGCAGTAGTGAGACCAGCCTCATTTGCTGTGTGTGTGTGTGTGTGTGTGTGT 759  
481 tcccttgaagagccttcaagagaaatacccaacagctgtgtaacaaacagctgtctc 540  
Db 760 TCCTCTGAAGAGACCTTTCAAGGAAAAATCCACACCTGTGTCAACAACAGCTGCT 819  
541 aagcagaacactgtct-gggatgctcctaanaacctgcagcaggaagaagatgaatac 599  
Db 820 AAGCAGAACACTGT 879  
600 aagctcccggtggtatcagtaactgtgtcctaaccattcagagagacaaatgtcgtc 659  
Db 880 AAGCTCCCGCGGATACTAGTAAGTGTCTAACAATTTTCAGAGAGAMAATGCTGCTC 939  
660 gtctct 665  
|||||  
Db 940 GTCTCT 945

RESULT 15  
BG260547 966 bp mRNA linear EST 13-FEB-2001  
LOCUS BG260547 602372096P1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4480366 5,  
DEFINITION mRNA sequence.  
ACCESSION BG260547  
VERSION BG260547.1 GI:12770363  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 966)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM10313 row: 0 column: 23  
 High quality sequence stop: 723.  
 Location/Qualifiers

## FEATURES

source

1.966  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4480366"  
 /clone\_1lb="NIH\_MGC\_93"  
 /issue\_type="transitional cell papilloma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dr primed.  
 Average insert size 1.7 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC library."

BASE COUNT 282 a 209 c 271 g 204 t  
 ORIGIN

Query Match 11.2%; Score 638.2; DB 10; Length 966;  
 Best Local Similarity 91.0%; Pred. No. 9,1e-102;

Matches 762; Conservative 0; Mismatches 53; Indels 22; Gaps 7;

OY 1774 ccacttaatgatcaggatctcagagccttggtgtcagtgtaacaggtacacggtcaaaa 1833  
 |||||||  
 Db 1 CCACCTTAATGATTCAGGATCTGAGCCTTGCTGTCAGTCAAGGTAAAGSTAAACGGTCAAAA 60  
 OY 1834 gagaaccagcagatttggaatcttctgcaagtcattatataatgaatgaagagcatct 1893  
 |||||||  
 Db 61 GAGAACCCAGCAGATTGGGAATCTTTGCAAGTCCATTATTAATGAGAGCAGCATCT 120  
 OY 1894 aaagatgaagagcttcgggtgaatgaatcaactgataagcagtaaatggaatccctgttg 1953  
 |||||||  
 Db 121 AAAGATGGAAGCGCTGCGGTGAATGATCAACGTATAGCAGTAAATGAGAGATCCCTGTTG 180  
 OY 1954 ggcaagacaacccaagaatgcatgtaaacccaagaaggtctatgctactgaagcaat 2013  
 |||||||  
 Db 181 GGCAGACAAACCAAGTGCATGAAACCCITAAAGAGCTATGCTACTGAGGCAAT 240  
 OY 2014 aaacgaggaatgatccagcttatgttgcaaggagaataagcaagtgcagtgcgaag 2073  
 |||||||  
 Db 241 AAACGAGAGATGATCCAGCTTATTTGTGCAAGAGATAAGCAAGTCAATGAGCTGAAG 300  
 OY 2074 tcaactggagcccccttgagacctgagcctgccaatgaacaacagcgttgatgaataga 2133  
 |||||||  
 Db 301 TCACCTGGGAGCCCCCTGGACCTGAGCTGCCATTGAAACGCTTGATGATAGAA 360  
 OY 2134 cgaagaattcccatctcctctacagtggatggaagggtctgataatgcgccagaga 2193  
 |||||||  
 Db 361 CGAAGAAATTTCCATCTCCCTACAGTGGATGAGGGGCTTGATGAATGCCCAGCAGA 420  
 OY 2194 aatgctgcctcagtaggataatggtgagtcagtaataacacagcgtcccccacagtg 2253  
 |||||||  
 Db 421 AATGCTGCCCTCAGTAGATATATG-----GGTAAATACAGCTGTCCCTACAGTG 471  
 OY 2254 aatatgcccaagaatgacactcatatataagaatatacaggttgccagtgtctctcca 2313  
 |||||||  
 Db 472 AATATGCCCAAGATGACATGTCATATATAGAGATGACAGTTCGCCAGTCTCTCA 531  
 OY 2314 catctctgaccagctctcttccagctcccatatgatgtgagggttgtagaggaagat 2373  
 |||||||  
 Db 532 CATCTCTCTGACCACTCTCTTCCAGCTCCCATGATGATGAGG-TTGTACGCGCAGAT 590  
 OY 2374 gctgtactggccaagcctgcaatcagtgatcagtcagcagctctcttgatccagat 2433  
 |||||||

Db 591 GCTGTAAGTGGGCCAA-GCTGCATCAGGTGATTCAGCCGACTGCTTTGAGTCCAGAT 649  
 OY 2434 gtgatccag--ttcttgctttcaagaagaagattggacgtcagagtaatgcaag--- 2488  
 |||||||  
 Db 650 GTAGATCCAGGCTTCTTGCTTTCACGAAGAGGAGATTGCCACGTCAGAACTATGGTC 709  
 OY 2489 ---aaaaacacaaag--caatttcagatgccaagcaatggatttggttaaacag 2543  
 |||||||  
 Db 710 AGAAAAACGCACAAAGCAGATTTCAGATGCCAGTCAATGG-GTTCGCTAAACACGA 768  
 OY 2544 aaatcaaaaaagcatggaattaggtatagctgacgaagactaaactcaataacagtgga 2600  
 |||||||  
 Db 769 AAGTTCAAAAGCTGAGATTAAAGTATAGCTGGAAGACTTAATCCAAAACGGGCTA 825

Search completed: July 24, 2002, 09:16:29  
 Job time: 9426 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 06:47:48 ; Search time 101.24 Seconds  
(without alignments) 13802.934 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689

Sequence: 1 atgaaagtacccgtgtgtctt.....tcttcctgttaaaaaaaaaa 5689

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCrus.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126	2.2	1042	4 US-09-276-531-77	Sequence 77, Appl
2	105.6	1.9	7218	4 US-08-232-463-14	Sequence 14, Appl
3	58.8	1.0	7431	4 US-09-306-998-2	Sequence 2, Appl
4	52.2	0.9	8040	1 US-08-536-281-1	Sequence 1, Appl
5	52.2	0.9	8040	3 US-09-100-804-1	Sequence 1, Appl
6	51.4	0.9	87350	3 US-08-781-891-79	Sequence 79, Appl
7	51	0.9	696	4 US-09-461-697-193	Sequence 193, App
8	51	0.9	699	4 US-09-461-697-191	Sequence 191, App
9	51	0.9	717	4 US-09-461-697-189	Sequence 189, App
10	51	0.9	774	4 US-09-461-697-187	Sequence 187, App
11	51	0.9	819	4 US-09-461-697-185	Sequence 185, App
12	51	0.9	1669	4 US-09-461-697-184	Sequence 184, App
13	50.6	0.9	8043	5 PCT-US94-09943-1	Sequence 1, Appl
14	50.6	0.9	8119	4 US-09-290-640-45	Sequence 45, Appl
15	49	0.9	5836	4 US-09-233-086-2	Sequence 2, Appl
16	47.8	0.8	11298	3 US-08-948-705-3	Sequence 3, Appl
17	47.8	0.8	11721	4 US-09-026-039-3	Sequence 3, Appl
18	47.8	0.8	11721	4 US-09-026-039-3	Sequence 3, Appl
19	47.2	0.8	376	2 US-08-623-906A-18	Sequence 18, Appl
20	46.8	0.8	2447	2 US-09-014-969-14	Sequence 14, Appl
21	46	0.8	5552	3 US-08-155-888-1	Sequence 1, Appl
22	45.8	0.8	176373	3 US-09-128-155-17	Sequence 17, Appl
23	45.6	0.8	712	1 US-08-410-804-4	Sequence 4, Appl
24	45.6	0.8	712	1 US-08-259-514-4	Sequence 4, Appl
25	45.6	0.8	712	2 US-08-858-311-4	Sequence 2, Appl
26	45.6	0.8	1830	1 US-08-410-804-2	Sequence 1, Appl
27	45.6	0.8	1830	1 US-08-259-514-2	Sequence 2, Appl

28	45.6	0.8	1830	2 US-08-858-311-2	Sequence 2, Appl
29	45.6	0.8	15144	3 US-08-458-434A-6	Sequence 6, Appl
30	45.6	0.8	73604	4 US-09-268-992-7	Sequence 7, Appl
31	45.2	0.8	306	4 US-09-461-697-203	Sequence 203, App
32	45.2	0.8	152331	3 US-09-128-155-16	Sequence 16, Appl
33	45	0.8	2223	1 US-08-257-073-4	Sequence 4, Appl
34	44.8	0.8	240	1 US-08-628-417-6	Sequence 6, Appl
35	44.6	0.8	289	4 US-09-007-005-17	Sequence 17, Appl
36	44.6	0.8	372	4 US-09-244-796-17	Sequence 17, Appl
37	44.6	0.8	829	4 US-09-018-584A-13	Sequence 13, Appl
38	44.6	0.8	6755	3 US-08-931-999-4	Sequence 4, Appl
39	44.4	0.8	5433	3 US-08-929-329-1	Sequence 1, Appl
40	44	0.8	494	2 US-08-332-766A-22	Sequence 22, Appl
41	43.4	0.8	1051	4 US-09-245-041-10	Sequence 10, Appl
42	43.4	0.8	12047	2 US-09-022-461-1	Sequence 1, Appl
43	43.2	0.8	8920	4 US-08-446-855A-1	Sequence 1, Appl
44	43.2	0.8	8920	4 US-09-150-741-1	Sequence 1, Appl
45	43.2	0.8	152331	3 US-09-128-155-16	Sequence 16, Appl

## ALIGNMENTS

RESULT 1  
US-09-276-531-77  
Sequence 77, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSER: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lynn E. Murty, Ph.D.  
REGISTRATION NUMBER: 42,918  
REFERENCE/DOCKET NUMBER: PA-0008 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 77:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1042 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BLADTUT04

CLONE: 1555118  
US-09-276-531-77

Query Match 2.28; Score 126; DB 4; Length 1042;  
Best Local Similarity 66.7%; Pred. No. 8.7e-26;  
Matches 212; Conservative 0; Mismatches 100; Indels 6; Gaps 2;

```
QY 2510 cagatgcacatcatttgatttgcgttaaacacgaacaaatcaaaagatgattagta 2569
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 CACGGCGAGGACCGCTGAAATCAATTAATTTGAAAGCCGCGAAGAGATGAGACT--TG 587
QY 2570 taqctgcagagactaacatcaatcagtgatgagccagaagcaggtctcccgacag 2629
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 588 TGGCAGATGAAGCAAGATTCACATCTGCTGAGCAAAAATCGAATCTCCAGCAAG 647
QY 2630 atgtgagctccctccctggtctgagaagaatcagctgttgagagatctgcagacg 2669
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 648 ATTTGGCTCAACTGTGGTTTGAAGAAAGTCACGCTCTTGAGAGAGTCTGACAGCTG 707
QY 2690 ttccagagtgacttgaatgaggatattccttcacatgctgcacgcccggaglaa 2749
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 708 TGGCGGAGGTCAGAGAAAT---GACCTTCCCTTCAAGGCCCGCGCATGTTTC 764
QY 2750 gagcgagggatgcaatgagagcttcagagctgcacatgcagaacatctatgataa 2809
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 765 GAGGCCAGAGCTGCAATGAGAGCTTTAGAGAGCATTGCAAAATCTAGAGTGACCTG 824
QY 2810 cagtagatgatgatg 2827
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 825 AAGAAATAGAGCTGACG 842
```

## RESULT 2

US-08-232-463-14/C  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 diagonal Road, Suite 500  
City: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZpT-Fls  
US-08-232-463-14

Query Match 1.98; Score 105.6; DB 1; Length 7218;  
Best Local Similarity 5.18; Pred. No. 2.2e-19;  
Matches 21; Conservative 264; Mismatches 123; Indels 0; Gaps 0;

```
QY 2825 atgaagcattgagacttgaagaacagcagaagaattcaagtcgggagagagt 2884
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1465 AAGTAGTTAAGATGAAAGATTTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRR 1406
QY 2885 ctgatacagcagcagtgatcagcttcacatcctctcgagagacaatgaaacc 2944
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1405 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1346
QY 2945 aagagaaagtgtataagactgtatgaaaaagataaactgtgaaaaagaagaag 3004
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1345 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1286
QY 3005 atagagataaggaagaagataaagcagaaggaagatctgaaggtctggag 3064
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1285 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1226
QY 3065 acatgtcaggttgatgcacacatcgaagaatgacaaatgagaacacggtataa 3124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1225 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1166
QY 3125 aaatacaggaatccttacatcagaagagagagatagatgaacagagcagagaga 3184
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1165 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1106
QY 3185 ggaltaagccaacactcgaaatttaggaacagacagctcgagac 3232
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1105 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1058
```

## RESULT 3

US-09-306-998-2  
Sequence 2, Application US/09306998  
Patent No. 6291173  
GENERAL INFORMATION:  
APPLICANT: Bartel, Paul L.  
APPLICANT: Tavligian, Sean V.  
TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein  
FILE REFERENCE: MMSC2  
CURRENT APPLICATION NUMBER: US/09/306,998  
CURRENT FILING DATE: 1999-05-07  
EARLIER APPLICATION NUMBER: US 60/084,740  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 7431  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (57)..(6167)  
US-09-306-998-2

Query Match 1.04; Score 58.8; DB 4; Length 7431;  
Best Local Similarity 52.4%; Pred. No. 5.2e-06;  
Matches 129; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 1383 tatcagcttaagaaggtacagaaggttgagatcagcatcctccagagatgtaac 1442



Db 5918 tattaacatgagcgagagcagatggtccttcagtagtctgagagatcgagcag 5977  
Qy 1443 aatagtgctcagctcaatcatalgtgaaaaacatctcccccggggggcccatlca 1502  
Db 5978 cccatctgagagactaccatcattatgttaaaacagtggttgaagaaggagcagctctga 6037  
Qy 1503 ggaatggccgaacttaaggaagagacagactatagaggttaaatggaatgattgtgg 1562  
Db 6038 agacggagctctgaaaaaggggcgatcagatcattgtctcattagggcagagctcaagg 6097  
Qy 1563 caatcccaagaagaagtgttgcctgtgagaagcaccagaatgagaaggaactgtgag 1622  
Db 6098 agtcaaccatgaagaagctgtgcatccatcccttaaacgacaaaggcaactgtcattgat 6157  
Qy 1623 ccttct 1628  
Db 6158 ggttct 6163

RESULT 4  
US-08-596-291-1  
: Sequence 1, Application US/08596291  
: Patent No. 5821075

GENERAL INFORMATION:  
APPLICANT: GONEZ, LEONEL JORGE  
APPLICANT: SARAS, JAN  
APPLICANT: CLAESSON-WELSH, LENA  
APPLICANT: HELDIN, CARL-HENRIK  
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASES  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/596,291  
FILING DATE: 09-AUG-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/115,573  
FILING DATE: 01-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: 10461/7000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
TELEX: 92-1742 EZEKIL  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8040 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: HOMO SAPIENS  
FEATURE:  
NAME/KEY: CDS

LOCATION: 78..7475  
US-08-596-291-1

Query Match 0.9%; Score 52.2; DB 1; Length 8040;  
Best Local Similarity 51.5%; Pred. No. 0.00042;  
Matches 120; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 1369 ggcagaaggttaataatccagcttaagaaggtaagaaggtttggatlcagcact 1428  
Db 4107 GGAGATATCTTTGAGGTTGACCTGACCTGTAATAAATGATTAACAGCTGGGCAATGTCACAG 4166  
Qy 1429 tccagagatgtacaatagttgttcagctccaatctatgtgaaaaaacattcccccgg 1488  
Db 4167 GGAGGTGTGAATATACAGTGTCAAGACATGCTGGCATTATATGTAAGATGTATTCGCCAG 4226  
Qy 1489 gggcgccatccagatgcccagcttaaggcagagagacagactataagtgaaatga 1548  
Db 4227 GGAGCAGCAGACTCTGATGTGTGAATTCACAAAGGATGATCGCTCTACCTGCAATGGA 4286  
Qy 1549 gtagatttagtggcaaatcccaagaagaagttgttcgctgttgagaagcac 1601  
Db 4287 GTTAGTCTAGAAAGAGGCCACCATTAAGCAAGCTGTGGAACACTGGAATATAC 4339

RESULT 5  
US-09-100-804-1  
: Sequence 1, Application US/09100804  
: Patent No. 6066472

GENERAL INFORMATION:  
APPLICANT: GONEZ, LEONEL JORGE  
APPLICANT: SARAS, JAN  
APPLICANT: CLAESSON-WELSH, LENA  
APPLICANT: HELDIN, CARL-HENRIK  
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASES  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/100,804  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/596,291  
FILING DATE: 09-AUG-1996  
APPLICATION NUMBER: US 08/115,573  
FILING DATE: 01-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/09943  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: 10461/7003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8040 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: HOMO SAPIENS  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 78..7475  
US-09-100-804-1

Query Match 0.9%; Score 52.2; DB 3; Length 8040;  
Best Local Similarity 51.5%; Pred. No. 0.00042;  
Matches 120; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 1369 gccaaggcctaataatccagcctaagaaggtacagaaggttgagcattcagcact 1428  
DB 4107 GCAGATATCTTTGAGGTTGAACGTGCTAAATATGAACACCTTGGGATTAAGTCAAG 4166  
QY 1429 tccagagtgtaacaataagtggtcagctccaatctatgtgaaaaacattccccc 1488  
DB 4167 GGAGGTGTAATACAGAGTGTCAAGACATGTGGCATTATGTGAAGAAGTATTCGCCAG 4226  
QY 1489 gggggggccatcagatggtccgacttaaggcagagagacagactatagagtaatgga 1548  
DB 4227 GGAGCAGCAGAGACTGTGAGTGAATTCACAAGGTGATCCGTCCTGATGCTCAATGGA 4286  
QY 1549 gtaagttagtgaggcaaatcccaagaggaagttgttcgcctgtgagaagcac 1601  
DB 4287 GTTACTTGAAGAGGCCACCCATTAAGCAAGCTGTGGAACACATGAGAAATAC 4339

## RESULT 6

US-08-781-891-79/c  
Sequence 79, Application US/08781891

PATENT NO. 6090620  
GENERAL INFORMATION:  
APPLICANT: Yu, Ying-Hui  
APPLICANT: Fu, Chang-Eu  
APPLICANT: Oshima, Junko  
APPLICANT: Mulligan, John T.  
APPLICANT: Schellenberg, Gerald D.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,891  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6090620tenburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052.419  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:

LENGTH: 87350 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-781-891-79

Query Match 0.9%; Score 51.4; DB 3; Length 87350;  
Best Local Similarity 47.6%; Pred. No. 0.0036;  
Matches 151; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 2920 ctggaagagcaaatgaatgaaacccaagaagaagtgatgaagactgtatagaataaagat 2979  
DB 55399 CTTGAAG 55340  
QY 2980 aaaaactggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 3039  
DB 55339 AAG 55220  
QY 3040 aagggaatgctgaaggtcttggaagacatgctcaggttctgcaaacatcgaaagaatgac 3099  
DB 55279 AAGCGAGGCGCGTCCGACGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 55220  
QY 3100 aagattgagaagaacggttaataataataacaggaatccttaacatcagaagaagaag 3159  
DB 55219 GAG 55160  
QY 3160 atcgaatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3219  
DB 55159 GAG 55100  
QY 3220 caagctcgagagcgtga 3236  
DB 55099 AAGCAG 55083

## RESULT 7

US-09-461-697-193  
Sequence 193, Application US/09461697

PATENT NO. 6277974  
GENERAL INFORMATION:  
APPLICANT: COGENET NEUROSCIENCE, Inc.  
APPLICANT: Lo, Donald C.  
APPLICANT: Barney, Shawn  
APPLICANT: Thomas, Mary Beth  
APPLICANT: Portbury, Stuart D.  
APPLICANT: Putnam, Kasturi  
APPLICANT: Katz, Lawrence C.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
FILE REFERENCE: 1001-005-999  
CURRENT APPLICATION NUMBER: US/09/461,697  
CURRENT FILING DATE: 1999-12-14  
NUMBER OF SEQ ID NOS: 466  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 193  
LENGTH: 696  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-461-697-193

Query Match 0.9%; Score 51; DB 4; Length 696;  
Best Local Similarity 46.4%; Pred. No. 0.00017;  
Matches 201; Conservative 0; Mismatches 230; Indels 2; Gaps 1;

QY 2752 ggcaggagatgcaatgagagctcagagctgcagcagcagcagcagcagcagcagcagcagc 2811  
DB 181 ggaagagaagaagaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 240  
QY 2812 gtaagtatgataagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2871

Db 241 gaagatgaaagatcaaaacgaagaagggagagcttgaaagaagacaaagttaa 300  
Qy 2872 tcaggagagagatctgtatccacagccagtgatcagccttccactctctggaagacaa 2931  
Db 301 aaggggagagagatggaagaagagataaaatgtgaatgagaagggagagatgcaaaa 360  
Qy 2932 atgaatgaaacaaagaaagtgatgaagctgtatgaaaaaagataaaacttgaaaa 2991  
Db 361 gagaagaagatggaaaaaagtg-aaagcgaagaagaatgtgagaagctggaag 418  
Qy 2992 gaaagaagaagatagagataaggaagaataaatgaaagccaaaggaaggaagctgctg 3051  
Db 419 agaaaggaagaagatgaaagaaggaagaagaagaagaacaggaagctggaagaaga 478  
Qy 3052 aaggccttggaagacatgtcaggtttgcaaacatcgaaagaatgacagaatgagaa 3111  
Db 479 atgaagatggaagaagagagagagataaaagagggaggaagctgtaaaagcagaag 538  
Qy 3112 acgggtataataaaatcagaagatcctttacatcagaagaagagagatacgaatgaag 3171  
Db 539 atgaaagaagaagagagagatggaagaagaatgaaagctggaagagagagagag 598  
Qy 3172 caggagcagagaga 3184  
Db 599 aaggaagaagaaga 611

RESULT 8  
US-09-461-697-191  
; Sequence 191, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: LO, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Putnam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 191  
; LENGTH: 699  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-461-697-191

Query Match 0.9%; Score 51; DB 4; Length 699;  
Best Local Similarity 46.4%; Pred. No. 0.00018;  
Matches 201; Conservative 0; Mismatches 230; Indels 2; Gaps 1;

Qy 2752 ggcagggatgcatgagagcttcagagctgcacatcgacaatcttatgataaaccgcg 2811  
Db 184 gggaggaagaagaagacagctgagcagaagaataaaatgaaagaagaatcagaaga 243  
Qy 2812 gtatagatgatatgaaagcagcttggaagaagacacagaagaaggttcaaga 2871  
Db 244 gaagatgaaagatgcaaaacgaagaagaagggagagcttggaagaagaagacaaatgaa 303  
Qy 2872 tcaggagagagctgtatccacagccagtgatcagccttccactctctgagagacaa 2931  
Db 304 aagggggaagaagatggaagaagagataaaatggaatgagaagaaggaagatgcaaaa 363  
Qy 2932 atgaatgaaacaaagaaaggtatgaagctgtatgaaaaaagataaaacttgaaaa 2991  
Db 364 gagaagaagatggaaaaaagtg-aaagcgaagaagaagaatggaagaagctggaag 421

Qy 2992 gaaagaagaagatagagataaggaagaagataaaatgaaagccaaagaaggaatgctg 3051  
Db 422 agaaaggaagaagatgtaaaagaggaagaagacagaagaagaacaggaagtggaagaaga 481  
Qy 3052 aaggccttggaagacatgtcaggtttgcaaacatcgaaagaatgacagaatgagaa 3111  
Db 482 atgaagatggaagaagaagagagataaaagaaggggaaagatgtaaagctcaagaag 541  
Qy 3112 acgggtataataaaatcagaagatcctttacatcagaagaagagagagatacgaatgaag 3171  
Db 542 atgaaagaagaagagagagatggaagaagaatgaaagatgaagctggaagaagcttgaa 601  
Qy 3172 caggagcagagaga 3184  
Db 602 aaggaagaagaaga 614

RESULT 9  
US-09-461-697-189  
; Sequence 189, Application US/09461697  
; Patent No. 6277974  
; GENERAL INFORMATION:  
; APPLICANT: COGENT NEUROSCIENCE, Inc.  
; APPLICANT: LO, Donald C.  
; APPLICANT: Barney, Shawn  
; APPLICANT: Thomas, Mary Beth  
; APPLICANT: Portbury, Stuart D.  
; APPLICANT: Putnam, Kasturi  
; APPLICANT: Katz, Lawrence C.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
; TITLE OF INVENTION: CELL DEATH  
; FILE REFERENCE: 10001-005-999  
; CURRENT APPLICATION NUMBER: US/09/461,697  
; CURRENT FILING DATE: 1999-12-14  
; NUMBER OF SEQ ID NOS: 466  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 189  
; LENGTH: 717  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-461-697-189

Query Match 0.9%; Score 51; DB 4; Length 717;  
Best Local Similarity 46.4%; Pred. No. 0.00018;  
Matches 201; Conservative 0; Mismatches 230; Indels 2; Gaps 1;

Qy 2752 ggcagggatgcatgagagcttcagagctgcacatcgacaatcttatgataaaccgcg 2811  
Db 202 gggaggaagaagaagacagcagtgagcagaagaataaaatgaaagaagaatcagaaga 261  
Qy 2812 gtatagatgatatgaaagcagcttggaagaagacacagaagaagaaggttcaaga 2871  
Db 262 gaagatgaaagatgcaaaacgaagaagaagggagagcttggaagaagaagacaaatgaa 321  
Qy 2872 tcaggagagagctgtatccacagccagtgatcagccttccactctctgagagacaa 2931  
Db 322 aagggggaagaagatggaagaagaagataaaatggaatgagaagaaggaagatgcaaaa 381  
Qy 2932 atgaatgaaacaaaggaagaagtgataagactgtatgaagaagaagataaaacttgaaaa 2991  
Db 382 gagaagaagatgtaaaagaagtg-aaagcgaagaagaaggaatggaagaagatggaagaag 439  
Qy 2992 gaaagaagaagaatagagataaggaagaagataaaatgaaagccaaagaaggaatgctg 3051  
Db 440 agaaaggaagaagatgaaagaagaggaagaagacagaagaagaacaggaagtgtaaaagaaga 499  
Qy 3052 aaggccttggaagacatgtcaggtttgcaaacatcgaaagaagatgacagaattgagaa 3111  
Db 500 atgaagatggaagaagaagagagataaaagaagggggaagaagatgtataaagccaagaag 559

QY 3112 acgggttaataataataacaggaatccttaacatcagaagagggagagatgaag 3171  
Db 560 atgaataaagagagagagagatggaataaagaagatgaagtggaatggaagagagcctggaa 619  
QY 3172 caggagcaggaga 3184  
Db 620 aagagaagaaga 632

## RESULT 10

US-09-461-697-187  
Sequence 187, Application US/09461697  
Patent No. 6277974  
GENERAL INFORMATION:  
APPLICANT: COGENET NEUROSCIENCE, Inc.  
APPLICANT: Lo, Donald C.  
APPLICANT: Barney, Shawn  
APPLICANT: Thomas, Mary Beth  
APPLICANT: Portbury, Stuart D.  
APPLICANT: Putnam, Kasturi  
APPLICANT: Katz, Lawrence C.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
FILE REFERENCE: 10001-005-999  
CURRENT APPLICATION NUMBER: US/09/461,697  
NUMBER OF SEQ ID NOS: 466  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 187  
LENGTH: 774  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-461-697-187

Query Match 0.9%; Score 51; DB 4; Length 774;  
Best Local Similarity 46.4%; Pred. No. 0.00019;  
Matches 201; Conservative 0; Mismatches 230; Indels 2; Gaps 1;

QY 2752 ggcaggagatgcatgagagcttcagagctgcacatcttatgataaaccgcg 2811  
Db 259 ggagggataaagaagaagcagtgagcagagaagtaaaatgagaagaagatccaaga 318  
QY 2812 gtgagatgatatgataagagcagtgagccttggaagaagacacagaagaagttcaaga 2871  
Db 319 gaagatgaaagatcaaaaacgaagaagagggagagcttgaaagaagacaaagatgaa 378  
QY 2872 tcaggagagagctctgtatccacagcagtgatcagccttcacactctctgggagagcaa 2931  
Db 379 aaaggagagagatggaagaagataaaatggaatggaagaagagatgcaaa 438  
QY 2932 atgagatggaaccaaagaaagtgatagatcagatgataaagaagataaactggaaga 2991  
Db 439 gggagaagaatggaataaagagtg--aagcggaaagaagaaatggaagaagtggaaga 496  
QY 2992 gaaagagagagatagagatagaagaagaataaatagaagaacgaagaagatgctg 3051  
Db 497 agaaagagagatgaaagaagagaaagacagaagaagaacagagatggaagaaga 556  
QY 3052 aagggtctggagacatgctcaggttgcgaacatcgaaagaagatgacagagttgaga 3111  
Db 557 atgagatggaagaagaagagagataaagaagaggggaagatgtaaaatgtaagaag 616  
QY 3112 acgggttaataataataacaggaatccttaacatcagaagagagagatgaag 3171  
Db 617 atgaataaagagagagatggaagaagaagatggaagtggaatggaagaagcctggaa 676  
QY 3172 caggagcaggaga 3184  
Db 677 aagagaagaaga 689

RESULT 11  
US-09-461-697-185  
Sequence 185, Application US/09461697  
Patent No. 6277974  
GENERAL INFORMATION:  
APPLICANT: COGENET NEUROSCIENCE, Inc.  
APPLICANT: Lo, Donald C.  
APPLICANT: Barney, Shawn  
APPLICANT: Thomas, Mary Beth  
APPLICANT: Portbury, Stuart D.  
APPLICANT: Putnam, Kasturi  
APPLICANT: Katz, Lawrence C.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
FILE REFERENCE: 10001-005-999  
CURRENT APPLICATION NUMBER: US/09/461,697  
NUMBER OF SEQ ID NOS: 466  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 185  
LENGTH: 819  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-461-697-185

Query Match 0.9%; Score 51; DB 4; Length 819;  
Best Local Similarity 46.4%; Pred. No. 0.00019;  
Matches 201; Conservative 0; Mismatches 230; Indels 2; Gaps 1;

QY 2752 ggcaggagatgcatgagagcttcagagctgcacatcttatgataaaccgcg 2811  
Db 304 ggagggataaagaagaagcagtgagcagagaagtaaaatgagaagaagatccaaga 363  
QY 2812 gtgagatgatatgataagagcagtgagccttggaagaagacacagaagaagttcaaga 2871  
Db 364 gaagatgaaagatcaaaaacgaagaagagggagagcttgaaagaagacaaagatgaa 423  
QY 2872 tcaggagagagctctgtatccacagcagtgatcagccttcacactctctgggagagcaa 2931  
Db 424 aaaggagagagatggaagaagataaataggaatggaagaagatgcaaa 483  
QY 2932 atgagatggaaccaaagaaagtgatagatcagatgataaagaagataaactggaaga 2991  
Db 484 gagaagaagaatggaataaagagtg--aagcggaaagaagaaatggaagaagtggaaga 541  
QY 2992 gaaagagagagatagagataagagataaagaagataaatagaagaacgaagaagatgctg 3051  
Db 542 agaaagagagatgaaagaagagaaagagagagacagaagaagaacaggaatggaagaaga 601  
QY 3052 aagggtctggagacatgctcaggttgcgaacatcgaaagaagatgacagagttgaga 3111  
Db 602 atgagatggaagaagaagagagataaagaagaggggaagatgtaaaatgtaagaag 661  
QY 3112 acgggttaataataataacaggaatccttaacatcagaagagagagatgaag 3171  
Db 662 atgaataaagagagagatggaagaagaagatggaagaagaagtggaatggaagaagcctggaa 721  
QY 3172 caggagcaggaga 3184  
Db 722 aagagaagaaga 734

RESULT 12  
US-09-461-697-184  
Sequence 184, Application US/09461697  
Patent No. 6277974  
GENERAL INFORMATION:  
APPLICANT: COGENET NEUROSCIENCE, Inc.  
APPLICANT: Lo, Donald C.  
APPLICANT: Barney, Shawn  
APPLICANT: Thomas, Mary Beth

APPLICANT: Portbury, Stuart D.  
APPLICANT: Putnam, Kasturi  
APPLICANT: Katz, Lawrence C.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
TITLE OF INVENTION: CELL DEATH  
FILE REFERENCE: 10001-005-999  
CURRENT APPLICATION NUMBER: US/09/461,697  
CURRENT FILING DATE: 1999-12-14  
NUMBER OF SEQ ID NOS: 466  
SOFTWARE: FastSeq for Windows version 4.0  
SEQ ID NO: 184  
LENGTH: 1669  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-461-697-184

Query Match 0.9%; Score 51; DB 4; Length 1669;  
Best Local Similarity 46.4%; Pred. No. 0.00032;  
Matches 201; Conservative 0; Mismatches 230; Indels 2; Gaps 1;

QY 2752 ggcagggagtcgacatgagcgttcacagcgttcacatgacacatcttatgataaccgcg 2811  
DB 383 ggaagagaagaagaagaagcagtcgacagcagaagtaaaaaatgaaagaagaagtcagaa 442  
QY 2812 gtagatgtagatgtagaagcagtcgacacgttcgaaagaagacacagaagaagttcaaga 2871  
DB 443 gaagatgaaagatcaaaaacgaagaagaagggaagcgtcgaaagaagaacagacaaagtga 502  
QY 2872 tcaaggagagagtcgtatccacagcagtcagccttcacactctctgagagacaa 2931  
DB 503 aaaggaggaaagatggaagaagaataaaatgaaatgaaagaagaagaagtcgcaaa 562  
QY 2932 atggaatggaagcacaagagaaggtgtaagactgtagaagaagaataaactggaana 2991  
DB 563 ggaagaagaagatggaaaaaaagttg--aagacggaagaagaagatggaagatggaaga 620  
QY 2992 gaaagaagaagaagatagagaataagaagaataaataaagccagaagaaggaatgctg 3051  
DB 621 aagaagaggaagatgaaaaaggaagaagaacgaagaagaagaagttggaagaaga 680  
QY 3052 aaggccttgagacatgctcaggtttgcaaacacatcgaaagaatgcaagatggaaa 3111  
DB 681 atgaagatggaagaagaagaggagataaaagaagggaagaagtgaagaagtaagaag 740  
QY 3112 acgggtaaataaaatacagaatccttatacagaagaagagaggttaagaag 3171  
DB 741 atgaaaaagaagaagaagatggaagaagaatggaagtggaagaagaagtcgga 800  
QY 3172 caggagcaggaga 3184  
DB 801 aagagaagaaga 813

RESULT 13  
PCT-US94-09943-1  
Sequence 1, Application PC/TUS9409943  
GENERAL INFORMATION:  
APPLICANT:  
APPLICANT:  
APPLICANT:  
APPLICANT:  
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASES  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/09943  
FILING DATE: 01-SEP-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/115,573  
FILING DATE: 01-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: TWOMEY, MICHAEL J.  
REGISTRATION NUMBER: P-38,349  
REFERENCE/DOCKET NUMBER: L0461/7000WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
TELEX: 92-1742 EZEKIEL  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8043 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: HOMO SAPIENS  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 78..7478  
PCT-US94-09943-1

Query Match 0.9%; Score 50.6; DB 5; Length 8043;  
Best Local Similarity 51.1%; Pred. No. 0.0012;  
Matches 119; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 1369 ggcagaagccttaatacagcgttaagaagaagtagacagaaggtttggatcagcact 1428  
DB 4110 GGAGATATCTTTGAGTTGAGTGGCTMAAAATGATTAACGCTTGGGATTAAGTGCACG 4169  
QY 1429 tccagagatgaacaataggtgctcagctccaatctatgaaaaaacatttcctcccg 1488  
DB 4170 GGAGGTGTGATATGAGATGTCAGACATGTGCGCTTTATGTGAAAGCTGTTATCCCCAG 4229  
QY 1489 gggcgccactcaagatgcccgaactaaagcaggaagacagactatagaagtaaatgga 1548  
DB 4230 GGAGCAGCAGACGTCGTATGTAGATTCACAAAGCTGATCGCGTCCATACGTCATGGA 4289  
QY 1549 gtagattagtgggcaaatcccaagagaagttgcttcgctgtgagaagcac 1601  
DB 4290 GTTAGTCTAGAGAGCACCACCATTAAGCAAGCTGTGAAACACTGAGAAATAC 4342

RESULT 14  
US-09-290-640-45  
Sequence 45, Application US/09290640  
Patent No. 6204055  
GENERAL INFORMATION:  
APPLICANT: Dean, Nicholas M.  
APPLICANT: Marcussen, Eric G.  
TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signalling  
FILE REFERENCE: ISPH-0351  
CURRENT APPLICATION NUMBER: US/09/290,640  
CURRENT FILING DATE: 1999-04-12  
NUMBER OF SEQ ID NOS: 85  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 45  
LENGTH: 8119

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64)..(7521)
; PUBLICATION INFORMATION:
; JOURNAL: FEBS Lett.
; VOLUME: 337
; ISSUE: 2
; PAGES: 200-206
; DATE: 1994-01-10
; DATABASE ACCESSION NUMBER: D21209/Genbank
; DATABASE ENTRY DATE: 1999-02-05
US-09-290-640-45

```

```

Query Match
Best Local Similarity 0.98; Score 50.6; DB 4; Length 8119;
Matches 119; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

```

```

QY 1369 ggcagaagccttaataatccagcttaagaaggtacagaaggttggatccagcact 1428
DB 4153 ggaatatacttggagcttgcagctgcctcaaaaatgatacagcttgggataagtgtaacg 4212
QY 1429 tccagaagatgtaacaatagctgtgctcagctccaatctatgtgaaaacattctcccccg 1488
DB 4213 ggaagtgtaatacgaagtcacagacatggtgcatltagtaaaagctgtatcccccag 4272
QY 1489 ggggcggccaltcagaatggcgcacttaagcgagagacagactatagaagtaaatgaa 1548
DB 4273 ggaacgacagagctgtagtaaatccaaaggtgacgcgtcctagctgctcaatgaa 4332
QY 1549 gtagatttagtggcaaatcccaagaaggaagttgttcgctgttgagaagcac 1601
DB 4333 gtagctcagaaggaagcaccataagcaagctgtggaacacagagaataac 4385

```

## RESULT 15

```

US-09-233-086-2
; Sequence 2, Application US/09233086
; Patent No. 6337192
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavliqian, Sean V.
; TITLE OF INVENTION: MMS1 - An MMAC1 Interacting Protein
; FILE REFERENCE: MMS1 Gene
; CURRENT APPLICATION NUMBER: US/09/233,086
; EARLIER FILING DATE: 1999-01-19
; EARLIER APPLICATION NUMBER: US 60/071,861
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5836
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (115)..(5757)
US-09-233-086-2

```

```

Query Match
Best Local Similarity 0.98; Score 49; DB 4; Length 5836;
Matches 124; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```

```

QY 1381 aatccagccttaagaaggttggatcagcactccagagatgta 1440
DB 5509 attactcttgagaaagcctcgaagccttgagggttagatgtgaggggttaagaa 5568
QY 1441 acaatagtggtcagcctcaatctatgaaacaattctcccccggggcgcccat 1500

```

```

DB 5569 agtcccactggagactgtccaatltagtaagactgtattgcaaaaggagacagctgca 5628
QY 1501 caggatggccgaacttaaggcagagacagactatagaagtaaatgaaatagtg 1560
DB 5629 gatcagggccgattaaacgaggggaltcagattttagctgttaatlpgcgaacccctgaa 5688
QY 1561 ggcnaatcccaagaggaagttgttcgctgttgagaagcaccagaatggaaggaactgtg 1620
DB 5689 ggtgtactaatgagcaagcagtcgcacattctaaacaccagagagggactgttaacctta 5748
QY 1621 agccttctg 1629
DB 5749 actgtgctg 5757

```

Search completed: July 24, 2002, 11:13:56  
Job time: 15968 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 08:04:43 ; Search time 547.23 Seconds

(without alignments)  
17849.033 Million cell updates/sec

Title: us-09-757-781-20

Perfect score: 5689  
Sequence: 1 atgaaagtgcgcgtgtgctt.....tcttccttcaaaaaaaaaa 5689

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq\_032802:\*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5087	89.4	5510	22	AAH73337
2	2476.8	43.5	2718	22	AAH18587
3	2476.8	43.5	2718	22	AAH78369
4	1695.8	29.8	2534	21	AAAD01182
5	1638.6	28.8	1800	22	AAH72793
6	842	14.8	867	22	AAAF91875
7	819.4	14.4	823	22	AAAS27320
8	819.4	14.4	823	22	AAH33594
9	549.8	9.7	556	22	AAH07109

10	516.2	9.1	597	22	AAH71480	Human cervical can
11	482.4	8.5	540	21	AAC79294	Human lung tumour-
12	482.4	8.5	540	23	AAAD23370	Human lung tumour-
13	465.4	8.2	510	22	AAAF68902	DNA encoding novel
14	415.4	7.3	425	22	AAAF67583	Novel human polyonu
15	384.4	6.8	564	22	AAH12111	Human CDNA clone (
16	357.4	6.3	504	22	AAH70055	Human cervical can
17	334.8	5.9	1262	22	AAAF68903	DNA encoding novel
18	282.6	5.0	305	22	AAH72113	Human cervical can
19	282.6	5.0	309	22	AAH71048	Human cervical can
20	271.6	4.8	339	22	AAAF69243	Human cervical can
21	253.8	4.5	400	22	AAAF66057	Novel human polyonu
22	253.4	4.5	582	23	AAAF68901	DNA encoding novel
23	182.8	3.2	556	22	AAAF61963	Human foetal liver
24	182.8	3.2	556	22	AAAF14889	Probe #10575 used
25	181	3.2	181	22	AAAF74458	Human foetal liver
26	181	3.2	181	22	AAAF54924	Probe #23610 used
27	173.8	3.1	180	21	AAAC02894	Human secreted pro
28	166	2.9	505	22	AAAF61506	Human foetal liver
29	166	2.9	505	22	AAAF29229	Probe #7695 for ge
30	166	2.9	505	22	AAAK09806	Human brain expres
31	166	2.9	505	22	AAAK35700	Human bone marrow
32	166	2.9	505	22	AAAF17106	Probe #7039 for ge
33	166	2.9	505	22	AAAF14115	Human cancer agent
34	162.8	2.9	2100	22	AAAF60861	Human cancer agent
35	162.8	2.9	2100	22	AAAF60956	Human cancer agent
36	162.8	2.9	2100	22	AAAF60963	Human cancer agent
37	162.8	2.9	2100	22	AAAF61007	Human cancer agent
38	160.4	2.8	1053	22	AAAF08393	Human PAR3 homolog
39	127.4	2.2	414	22	AAAF83888	Human poly nucleoti
40	126	2.2	1042	22	AAAF80367	Receptor #55 parti
41	118.8	2.1	565	23	AAAF68814	DNA encoding novel
42	112.2	2.0	145	21	AAAF45600	Human secreted exp
43	101.8	1.8	5407	23	ABF08765	Drosophila melanog
44	101.8	1.8	36336	23	ABF08764	Drosophila melanog
45	91.2	1.6	525	23	AAAF67317	DNA encoding novel

#### ALIGNMENTS

RESULT	1	
AAH73337		
AAH73337	standard; cDNA; 5510 BP.	
XX		
XX	AAH73337;	
XX		
XX	19-SEP-2001 (first entry)	
DE		
XX	Human cervical cancer marker nucleic acid 4611.	
DE		
KW	Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200142467-A2.	
XX		
PD	14-JUN-2001.	
XX		
XX		
PF	08-DEC-2000; 2000WO-US33312.	
XX		
PR	08-DEC-1999; 9905-0169681.	
PR	21-DEC-1999; 9905-0171350.	
PR	14-MAR-2000; 2000US-0189315.	
PR	12-MAY-2000; 2000US-0203791.	
PR	09-JUN-2000; 2000US-0210600.	
PR	21-JUL-2000; 2000US-0220114.	
XX		
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.	
XX		
PI	Schlegel R, Deeds J, Berger A, Zhao X;	
XX		
DR	WPI; 2001-375006/39.	

XX New isolated nucleic acid for diagnosing and treating cervical cancer  
PT and for assessing and detecting compounds for treating the cancer -  
XX  
PS Claim 1; Page 1029-1031; 1051P; English.  
XX The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.

XX  
Sequence 5510 BP; 1526 A; 1242 C; 1418 G; 1305 T; 19 other;

Query Match 89.4%; Score 5087; DB 22; Length 5510;  
Best Local Similarity 95.1%; Pred. No. 0;  
Matches 5411; Conservative 0; Mismatches 5; Indels 271; Gaps 5;  
QY 3 gaaagtgacggtgtgtcttgagacgacccgggtgtgtctgtccgttgcgggagcgccacat 62  
Db 3 gaaagtgacggtgtgtcttgagacgacccgggtgtgtctgtccgttgcgggagcgccacat 62  
QY 63 gaaagtttcaagcctcatccagcagcggttgacccggtcccggaagcgcatcgccaagga 122  
Db 63 gaaagtttcaagcctcatccagcagcggttgacccggtcccggaagcgcatcgccaagga 122  
QY 123 tccaaactcttgatacaggtgcatcgtcttgaaacatggaatggaatggaatactagact 182  
Db 123 tccaaactcttgatacaggtgcatcgtcttgaaacatggaatggaatggaatactagact 182  
QY 183 tgatgacatcttcttgatgtagcagacgataaagacagactgtagcagtggttgatga 242  
Db 183 tgatgacatcttcttgatgtagcagacgataaagacagactgtagcagtggttgatga 242  
QY 243 ggaagatccacatccacgaggtgtagtgcacagctgcagttccacgggttacccagagcc 302  
Db 243 ggaagatccacatccacgaggtgtagtgcacagctgcagttccacgggttacccagagcc 302  
QY 303 aagaatatttgtagtgggttgagcacaacaatgttccagcttccagcttccaaagc 362  
Db 303 aagaatatttgtagtgggttgagcacaacaatgttccagcttccagcttccaaagc 362  
QY 363 aacaagtgaatctgaggtgcacacacttcagctccttcgacaaatagctcttccttcg 422  
Db 363 aacaagtgaatctgaggtgcacacacttcagctccttcgacaaatagctcttccttcg 422  
QY 423 acgagtagtgacccagcttaatttgccctccaactctgtcagtgatagtaatttttc 482  
Db 423 acgagtagtgacccagcttaatttgccctccaactctgtcagtgatagtaatttttc 482  
QY 483 ctctgaagagcttccaagaaatacccaacgctgtgtccaacaacagctgtcttcctcaa 542  
Db 483 ctctgaagagcttccaagaaatacccaacgctgtgtccaacaacagctgtcttcctcaa 542  
QY 543 gcaagaacaactgtctgaggtctcctaacaactgcgacaggaagaatgaaactacagaag 602  
Db 543 gcaagaacaactgtctgaggtctcctaacaactgcgacaggaagaatgaaactacagaag 602  
QY 603 cctccgcggtatctagtaactgtcttaacaatttcagagagacaatgtcgtcgtc 662  
Db 603 cctccgcggtatctagtaactgtcttaacaatttcagagagacaatgtcgtcgtc 662  
QY 663 tctgagtgcagtgcaaccaatgtgtggaagtggtctgtagagaacaagaacagatgagga 722  
Db 663 tctgagtgcagtgcaaccaatgtgtggaagtggtctgtagagaacaagaacagatgagga 722  
QY 581 -----aggatgagga 590  
Db 581 -----aggatgagga 590  
QY 723 tgggacagagagatatacagtcgtgttgaaactgttggaacatgctgacaacgggtttgga 782  
Db 591 tgggacagagagatatacagtcgtgttgaaactgttggaacatgctgacaacgggtttgga 782  
591 tgggacagagagatatacagtcgtgttgaaactgttggaacatgctgacaacgggtttgga 650

QY 783 gcatatacccaacttctctctgtagatagtgtaaagctcgttagaagctcccaacgtag 842  
Db 651 gcatatacccaacttctctctgtagatagtgtaaagctcgttagaagctcccaacgtag 842  
QY 843 agggcctcgggaaatccatgtagtgccttcaagtgctcgtggaagcgcgcaaaccttgggtt 902  
Db 711 agggcctcgggaaatccatgtagtgccttcaagtgctcgtggaagcgcgcaaaccttgggtt 902  
QY 903 attagtaaacagattggaagaaagtgtgtaaaagcttgaaacatgaaactcttctgtgaa 962  
Db 771 attagtaaacagattggaagaaagtgtgtaaaagcttgaaacatgaaactcttctgtgaa 830  
QY 963 tgattgcatcttcagatgataatgtgtgcaccttcgaaataagaagattgaaacaaga 1022  
Db 831 tgattgcatcttcagatgataatgtgtgcaccttcgaaataagaagattgaaacaaga 890  
QY 1023 acatagtttcggcaagacatgctgtacacccatcatattgtgtcattgttctctgagc 1082  
Db 891 acatagtttcggcaagacatgctgtacacccatcatattgtgtcattgttctctgagc 950  
QY 1083 aataaagaagagatgataaacaactatcccaagtgagaagaacaattactattcaagcgg 1142  
Db 951 aataaagaagagatgataaacaactatcccaagtgagaagaacaattactattcaagcgg 1010  
QY 1143 ttttagccctacagccaagtatatgtgcaacaggaagtgtgaaagtgtagaggttcaac 1202  
Db 1011 ttttagccctacagccaagtatatgtgcaacaggaagtgtgaaagtgtagaggttcaac 1070  
QY 1203 ggtgcagagagaccccgactgaaaccccgcttgagcagatagactctcaccagact 1262  
Db 1071 ggtgcagagagaccccgactgaaaccccgcttgagcagatagactctcaccagact 1130  
QY 1263 accctatagcgcacaacccctcgggaaacccaacatccgctccagctcgtgcacctcaga 1322  
Db 1131 accctatagcgcacaacccctcgggaaacccaacatccgctccagctcgtgcacctcaga 1190  
QY 1323 tgtattatgtagcagctgtgaaagcagtggttaaacacaaanaaataagcgaaggtctta 1382  
Db 1191 tgtattatgtagcagctgtgaaagcagtggttaaacacaaanaaataagcgaaggtctta 1250  
QY 1383 tatccagcttaagaaggtgtagaaggtttgttgattccagatcaactccagagatgtaac 1442  
Db 1251 tatccagcttaagaaggtgtagaaggtttgttgattccagatcaactccagagatgtaac 1310  
QY 1443 aataagtggtcagctccaactctatgtgaaacaaatctcccccggggggcgccattca 1502  
Db 1311 aataagtggtcagctccaactctatgtgaaacaaatctcccccggggggcgccattca 1370  
QY 1503 ggtatgcggaacttaagcagcagcagcagacttatagaggttaaatgtagatlttagtg 1562  
Db 1371 ggtatgcggaacttaagcagcagcagcagacttatagaggttaaatgtagatlttagtg 1430  
QY 1563 caaatcccaagcgaagtggttctgtgtggaagcaccacgaatggaaggaagcgtgag 1622  
Db 1431 caaatcccaagcgaagtggttctgtgtggaagcaccacgaatggaaggaagcgtgag 1490  
QY 1623 ccttctggtccttcgcaggaagacgcttccacccaaggggaactgacagagcaag 1682  
Db 1491 ccttctggtccttcgcaggaagacgcttccacccaaggggaact- 1555  
QY 1683 ccagatgcagattcccaagaagcagaagcagaatgtagatatgttcttcaacctga 1742  
Db 1536 -----gaaagcagaagaatgtagatatgttcttcaacctga 1571  
QY 1743 tggcaccaggggaattcttgacattgaaagtcocacttaatatccagatctgagcgt 1802  
Db 1572 tggcaccaggggaattcttgacattgaaagtcocacttaatatccagatctgagcgt 1631  
QY 1803 tgggtcagtgatcaaaagtgtaacgggttcaaaaagaaacacgacagatttggaatcttgt 1862  
Db 1632 tgggtcagtgatcaaaagtgtaacgggttcaaaaagaaacacgacagatttggaatcttgt 1691



QY 1863 caagtcacatttaattgagagcagcaltcaagaatggaagcttcggtgtaattatca 1922  
|||||  
Db 1692 caaattccatttaattgagagcagcaltcaagaatggaagcttcggtgtaattatca 1751  
QY 1923 actatagaacataatgagaatccctgttggcagaacaacgaagtgcattggaac 1982  
|||||  
Db 1752 actatagaacataatgagaatccctgttggcagaacaacgaagtgcattggaac 1811  
QY 1983 cctaagaagcgtatgttactatgaaagcaataaaggaatgataccggttatgttc 2042  
|||||  
Db 1812 cctaagaagcgtatgttactatgaaagcaataaaggaatgataccggttatgttc 1871  
QY 2043 aagagaataaagcaagtcaatgagctgaagtcacctggagcccccctggaacctgaact 2102  
|||||  
Db 1872 aagagaataaagcaagtcaatgagctgaagtcacctggagcccccctggaacctgaact 1931  
QY 2103 gcccaattgaacacagcgttgggtgataagaagaattcccaattccctctaagtcg 2162  
|||||  
Db 1932 gcccaattgaacacagcgttgggtgataagaagaattcccaattccctctaagtcg 1991  
QY 2163 gattgaggggtctgataatcgcccaagaataatgtcctccatgataatgtgtga 2222  
|||||  
Db 1992 gattgaggggtctgataatcgcccaagaataatgtcctccatgataatgtgtga 2046  
QY 2223 gtcaagtaataacagcgttcccttaacagtgaatgtcccaagaatgacactgtcataat 2282  
|||||  
Db 2047 ---gttaataatacagcgttcccttaacagtgaatgtcccaagaatgacactgtcataat 2102  
QY 2283 aagaataatacagcgttcccttaacagtgtcctccatcctcttaacagcttcccttaacagctc 2342  
|||||  
Db 2103 aagaataatacagcgttcccttaacagtgtcctccatcctcttaacagcttcccttaacagctc 2162  
QY 2343 ccatgataatgttgggttctgtgaacgagatgtgtaatttggcagaagcgttcgaatcac 2402  
|||||  
Db 2163 ccatgataatgttgggttctgtgaacgagatgtgtaatttggcagaagcgttcgaatcac 2222  
QY 2403 tgattcagcagcagcgtccttctgaatcagaatgtgataccagcttctgttccaagaga 2462  
|||||  
Db 2223 tgattcagcagcagcgtccttctgaatcagaatgtgataccagcttctgttccaagaga 2282  
QY 2463 aggatttgaagctcagatgatagtcagaataacgcaagaacatttccaagtcagatca 2522  
|||||  
Db 2283 aggatttgaagctcagatgatagtcagaataacgcaagaacatttccaagtcagatca 2298  
QY 2523 attggaattcgttaaaaaacagaaaaatacaaaaagcattgattgaattagctgaacgagac 2582  
|||||  
Db 2299 -----attggaattcgttaaaaaacagaaaaatacaaaaagcattgattgaattagctgaacgagac 2312  
QY 2583 taaactcaatacagtgatgacccaagaagcaggttctccacagcagaatgtggttccttc 2642  
|||||  
Db 2313 taaactcaatacagtgatgacccaagaagcaggttctccacagcagaatgtggttccttc 2372  
QY 2643 cctgggtctgagaagaatgaagctgttggagagcttgcagaacccagatgtgcgaagtgcac 2702  
|||||  
Db 2373 cctgggtctgagaagaatgaagctgttggagagcttgcagaacccagatgtgcgaagtgcac 2432  
QY 2703 ttggaatggggaatatctcttccatcgttccacgagcggcggataatcagaagcaggagatg 2762  
|||||  
Db 2433 ttggaatggggaatatctcttccatcgttccacgagcggcggataatcagaagcaggagatg 2492  
QY 2763 caatgaagcgttccaagcttgcatacgaataatctatatgataaaacccggtgataatga 2822  
|||||  
Db 2493 caatgaagcgttccaagcttgcatacgaataatctatatgataaaacccggtgataatga 2552  
QY 2823 tgatgaagcgttgaagcgttgaagaagacacagaagaatgtcaagaatcagggagaga 2882  
|||||  
Db 2553 tgatgaagcgttgaagcgttgaagaagacacagaagaatgtcaagaatcagggagaga 2612  
QY 2883 gtctgtatccacagcagtgatcagcttccacactcctgagagacaaatgaaatgaa 2942  
|||||  
Db 2613 gtctgtatccacagcagtgatcagcttccacactcctgagagacaaatgaaatgaa 2672  
QY 2943 ccaagaagaagtgataagcctgatagaataaagataaaactgtaaaagaagaaga 3002

Db 2673 ccaagaagaagtgataagcctgatagaataaagataaaactggaataaagaagaaga 2732  
QY 3003 agatagaataaggaaggaatgaataatgaagccagaaggggaatctcaagggcttgg 3062  
Db 2733 agatagaataaggaaggaatgaataatgaagccagaaggggaatctcaagggcttgg 2792  
QY 3063 agacatgttcaagcttggcaaaacatcgaaaagaatgacaagatttgaataaacggttaaat 3122  
Db 2793 agacatgttcaagcttggcaaaacatcgaaaagaatgacaagatttgaataaacggttaaat 2852  
QY 3123 aaaaatacagaataatccttacaatcagaagaaggaatgacaatgaaacaggaacaga 3182  
Db 2853 aaaaatacagaataatccttacaatcagaagaaggaatgacaatgaaacaggaacaga 2912  
QY 3183 gaggattcaagccaaaacccaggaatttgaagaacagcagcttgaagcgtgataatgc 3242  
Db 2913 gaggattcaagccaaaacccaggaatttgaagaacagcagcttgaagcgtgataatgc 2972  
QY 3243 tgaataatcaagaatttccatcagacatttggcgtgtatgataatgtatgagggagat 3302  
Db 2973 tgaataatcaagaatttccatcagacatttggcgtgtatgataatgtatgagggagat 3032  
QY 3303 ttcttcttatgaaggttccatgtcctcaacgtctagaacctcagagccacgagaagagca 3362  
Db 3033 ttcttcttatgaaggttccatgtcctcaacgtctagaacctcagagccacgagaagagca 3092  
QY 3363 tatgataagcttctgtatgagccaaagtcaagaagccgggaattccaacccctcaactt 3422  
Db 3093 tatgataagcttctgtatgagccaaagtcaagaagccgggaattccaacccctcaactt 3152  
QY 3423 agacagtaacagatacaactccatgaacatcatgatcggataacggtctgaggaagaat 3482  
Db 3153 agacagtaacagatacaactccatgaacatcatgatcggataacggtctgaggaagaat 3212  
QY 3483 tcaagaagaagaagcaagaatgaaatgtgaagaatcgtcggagacataatgttgaagca 3542  
Db 3213 tcaagaagaagaagcaagaatgaaatgtgaagaatcgtcggagacataatgttgaagca 3272  
QY 3543 acccctggcaggaacgagcagcggcagcagcagcagcagcagcagcagcagcagcagcagc 3602  
Db 3273 acccctggcaggaacgagcagcggcagcagcagcagcagcagcagcagcagcagcagcagc 3332  
QY 3603 gcaagatgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3662  
Db 3333 gcaagatgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3392  
QY 3663 cagctctcgtcctcggcagaagcagaagaataatgcacagctcgttcccaagactcttggga 3722  
Db 3393 cagctctcgtcctcggcagaagcagaagaataatgcacagctcgttcccaagactcttggga 3452  
QY 3723 gcaagaactatccctcggggaagcgttccagaatgtgcagaagaagaacccagatcagcag 3782  
Db 3453 gcaagaactatccctcggggaagcgttccagaatgtgcagaagaagaacccagatcagcag 3512  
QY 3783 ctacaaggtccacagaagcagcgtacccctggagagacatggtcttcaacgcaaggtcatgct 3842  
Db 3513 ctacaaggtccacagaagcagcgtacccctggagagacatggtcttcaacgcaaggtcatgct 3572  
QY 3843 ggaataatcaggaagcttccctcggcagaagaacagagcgggaagggcagcaatgaagaagca 3902  
Db 3573 ggaataatcaggaagcttccctcggcagaagaacagagcgggaagggcagcaatgaagaagca 3632  
QY 3903 gctccttcggagggcccaagcaactatgactgtaataagaagtcaggaacccagttca 3962  
Db 3633 gctccttcggagggcccaagcaactatgactgtaataagaagtcaggaacccagttca 3692  
QY 3963 ggcctcccaaggggaccttccggcaagaatgtgccccctcccttccaaggttgcag 4022  
Db 3693 ggcctcccaaggggaccttccggcaagaatgtgccccctcccttccaaggttgcag 3752  
QY 4023 gctgaacagacttccagactcctgagaagaaggggaccttcttatctctgagcagcaata 4082

Db	3753	gctgaacagacttcaagactcctgtgnaaaggaggcccttctatcttcctgtgagacgcaata	3812
Qy	4083	acggaatgcttatgctgcgcgaataaagacatttccatataagactgtatatttggag	4142
Db	3813	acgagtgcttatgctgcgcgaataaagacatttccatataagactgtatatttggag	3872
Qy	4143	tttttttaaaccttcgataggtactactatgagatattctgtgtgtgtatcagtgctttaa	4202
Db	3873	tttttttaaaaccccgatgactactatgagagatttctgtgtgtgtatcagtgctttaa	3932
Qy	4203	gcggatgtatgagaaagaaatgnaaagccataatgtcttgcgcacataatccaaatgtctg	4262
Db	3933	gcggatgtatgagaaagaaatgnaaagccataatgtcttgcgcacataatccaaatgtctg	3992
Qy	4263	tttaatagaagatttcccaacccctgtgacaatacactgtgtttttagtgttccatagtctg	4322
Db	3993	tttaatagaagatttcccaacccctgtgacaatacactgtgtttttagtgttccatagtctg	4052
Qy	4323	cgctctccacagatgacaggaatctgcgcctactactatgattgtgtccgcgctgtgtg	4382
Db	4053	cgctctccacagatgacaggaatctgcgcctactactatgattgtgtccgcgctgtgtg	4112
Qy	4383	aaactccctgcacacacttgcagatgacaatccactgaagtggagagatgacgtgcgccc	4442
Db	4113	aaactccctgcacacacttgcagatgacaatccactgaagtggagagatgacgtgcgccc	4172
Qy	4443	tgattccccccttaagttcttccttaagctatgagatgagatcatcttgctctggtgagaa	4502
Db	4173	tgattccccccttaagttcttccttaagctatgagatgagatcatcttgctctggtgagaa	4232
Qy	4503	agtggtgcgacatacaccnaaaatltgggcttctgttacttcaacagcacagccattgctc	4562
Db	4233	agtggtgcgacatacaccnaaaatltgggcttctgttacttcaacagcacagccattgctc	4292
Qy	4563	gtactttgatacactgtgtgttcttccttcccttccagctcttgtgacgggagaatgc	4622
Db	4293	gtactttgatacactgtgtgttcttccttcccttccagctcttgtgacgggagaatgc	4352
Qy	4623	gtgtcaactatatacagaaggtlaagcgcatagtccaaacatgttgtgcacacatggggctcc	4682
Db	4353	gtgtcaactatatacagaaggtlaagcgcatagtccaaacatgttgtgcacacatggggctcc	4412
Qy	4683	ttttgttaacgctttatgtactaacaacttaacaataaagatgataccctgtgctgcgtta	4742
Db	4413	ttttgttaacgctttatgtactaacaacttaacaataaagatgataccctgtgctgcgtta	4472
Qy	4743	tacatacgcttgttcgctctgttccctgcacagttggtgtgataccaaagctctgtgtg	4802
Db	4473	tacatacgcttgttcgctctgttccctgcacagttggtgtgataccaaagctctgtgtg	4532
Qy	4803	gggaaagttgggagacagggatggctctctgcgggggggaagcttgggctgcatttggccctg	4862
Db	4533	gggaaagttgggagacagggatggctctctgcgggggggaagcttgggctgcatttggccctg	4592
Qy	4863	tgtctatctaggggagagcgaataaagaatattctccatagaagaagtcatggcccagt	4922
Db	4593	tgtctatctaggggagagcgaataaagaatattctccatagaagaagtcatggcccagt	4652
Qy	4923	acaatccatgaattatttaattagtttgttctgcacagcttgcacagaagggcacaga	4982
Db	4653	acaatccatgaattatttaattagtttgttctgcacagcttgcacagaagggcacaga	4712
Qy	4983	atgggacagaagataaacacagacagtcatttgcattgtcgtctctaacggtttttcaagtca	5041
Db	4713	atgggacagaagataaacacagacagtcatttgcattgtcgtcgtctctaacggtttttcaagtca	4772
Qy	5042	gaggaatgtatgctgtctaatgatcatcacacactgatgtctgcactgcgagtggccag	5101
Db	4773	gaggaatgtatgctgtctaatgatcatcacacactgatgtctgcactgcgagtggccag	4832
Qy	5102	cttccctaaagtattctgcatagaacataaagaacaaaggaaagccgtttacacatacac	5161
Db	4833	cttccctaaagtattctgcatagaacataaagaacaaaggaaagccgtttacacatacac	4892

QY	5162	agaagaaacatttcgcggtcccaacagcgtgtgtgtgcaggaagaactcactctcgtctcagat	5221
Db	4893	agaagaaacatttcgcggtcccaacagcgtgtgtgtgcaggaagaactcactctcgtctcagat	4952
QY	5222	agaagtcgtgtgtgtctcgcgggagatcccggtgtgtgtcccaactctctcttaattgttcgaa	5281
Db	4953	agaagtcgtgtgtgtctcgcgggagatcccggtgtgtgtcccaactctctcttaattgttcgaa	5012
QY	5282	catctctgattgttaacacatgtgctcggggtctaaagtgcctgtgaatcccgatgtgaaa	5341
Db	5013	catctctgattgttaacacatgtgctcggggtctaaagtgcctgtgaatcccgatgtgaaa	5072
QY	5342	aagctcgtgtgtgtgaagctcagcatatcatcttactttaaacaagaaaaaagaat	5401
Db	5073	aagctcgtgtgtgtgaagctcagcatatcatcttactttaaacaagaaaaaagaat	5132
QY	5402	gaatgtgatatctcatctttttttttatgtgcacattgtatcttctgtgtgaactgtttt	5461
Db	5133	gaatgtgatatctcatctttttttttatgtgcacattgtatcttctgtgtgaactgtttt	5192
QY	5462	agaagaatgtgtgtccacaacagataccggtctcctctcattctgtgtgaatgtctg	5521
Db	5193	agaagaatgtgtgtccacaacagataccggtctcctctcattctgtgtgaatgtctg	5252
QY	5522	ttctcttaacacgtgcggcggtgtctaaagtgtgttaccagatgtacgcgcagtgacttg	5581
Db	5253	ttctcttaacacgtgcggcggtgtctaaagtgtgttaccagatgtacgcgcagtgacttg	5312
QY	5582	gatgcacatgtgctcttctctacacgcctcccgatgtgtgaanaacagcttctctgtac	5641
Db	5313	gatgcacatgtgctcttctctacacgcctcccgatgtgtgaanaacagcttctctgtac	5372
QY	5642	ataatgaacactcctcaataaagaagcatattctctcgtttaaaaaaaa	5688
Db	5373	ataatgaacactcctcaataaagaagcatattctctcgtttaaaaaaaa	5419

```

RESULT 2
AAH18587
ID AAH18587 standard; cDNA; 2718 BP.
XX
AC AAH18587;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:18774.
XX
Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection

```



QY 2054 gcaagtcgaatgagctgaagtcacactgaggagcccccctggacctgagctgcccattgaa 2113  
 DB 1681 gcaagtcgaatgagctgaagtcacactgaggagcccccctggacctgagctgcccattgaa 1740  
 QY 2114 cagcgttggatgataagagaacgaagaatttcccatccctctacagtgaggattgaagggc 2173  
 DB 1741 cagcgttggatgataagagaacgaagaatttcccatccctctacagtgaggattgaagggc 1800  
 QY 2174 ttgatgaatcgcccaagcaagaatgctgcccctcagtaggaataatg 2233  
 DB 1801 ttgatgaatcgcccaagcaagaatgctgcccctcagtaggaataatg 1851  
 QY 2234 accaagcttccctcagtaggaataatgctgcccctcagtaggaataatg 2293  
 DB 1852 accaagcttccctcagtaggaataatgctgcccctcagtaggaataatg 1911  
 QY 2294 ggttgcagtagcttccctcagtaggaataatgctgcccctcagtaggaataatg 2353  
 DB 1912 ggttgcagtagcttccctcagtaggaataatgctgcccctcagtaggaataatg 1971  
 QY 2354 tggggttctgagcagcagatgctgtagcttggcgaagctgcaatcagtgatcagcag 2413  
 DB 1972 tggggttctgagcagcagatgctgtagcttggcgaagctgcaatcagtgatcagcag 2031  
 QY 2414 actgctcttggatcgaagatgctgtagcagcttctgcttccaaagagaagatttgagc 2473  
 DB 2032 actgctcttggatcgaagatgctgtagcagcttctgcttccaaagagaagatttgagc 2091  
 QY 2474 gtcagagtagtcgaagaacagcaagaatlttcagatgccaagtcgaatggatttcg 2533  
 DB 2092 gtcaga 2097  
 QY 2534 ttaaacacagaacaaacagcagtgattagtagtcgacgagactaaactcaata 2593  
 DB 2098 -----ctgacgagactaaactcaata 2118  
 QY 2594 cagtgatgacacagaagcaggttctcccaagagaatgctgagcttccctcagtgctga 2653  
 DB 2119 cagtgatgacacagaagcaggttctcccaagagaatgctgagcttccctcagtgctga 2178  
 QY 2654 agaaagtcagctcgttggagagcttgcagacgcagctgagcgaagtgacttgaatggg 2713  
 DB 2179 agaaagtcagctcgttggagagcttgcagacgcagctgagcgaagtgacttgaatggg 2238  
 QY 2714 atattccttccatcgctcagcgcgcggaataatcagaagcagggatgagcagtgagct 2773  
 DB 2239 atattccttccatcgctcagcgcgcggaataatcagaagcagggatgagcagtgagct 2298  
 QY 2774 tcagagctgacacagaacatcttatgataaacccgcgtagatgataatgataagga 2833  
 DB 2299 tcagagctgacacagaacatcttatgataaacccgcgtagatgataatgataagga 2358  
 QY 2834 tcagagctgacacagaacatcttatgataaacccgcgtagatgataatgataagga 2893  
 DB 2359 tcagagctgacacagaacatcttatgataaacccgcgtagatgataatgataagga 2418  
 QY 2894 cagcagtgatcagccttccatctctgagagacaaatgtaatgtaaacccaagaagaaag 2953  
 DB 2419 cagcagtgatcagccttccatctctgagagacaaatgtaatgtaaacccaagaagaaag 2478  
 QY 2954 gtcataagactgataaagaaagataaaacttgataaagaaagaaagataagata 3013  
 DB 2479 gtcataagactgataaagaaagataaaacttgataaagaaagaaagataagata 2538  
 QY 3014 aggaagaagataaagaaagcaagaagaaagaaatgctgaaagggcttgggagaaatgttca 3073  
 DB 2539 aggaagaagataaagaaagcaagaagaaagaaatgctgaaagggcttgggagaaatgttca 2598  
 QY 3074 ggttggcaaac 3085  
 DB 2599 ggcctggcaaac 2610

RESULT 3  
 AAH78369  
 ID AAH78369 standard; DNA; 2718 BP.  
 AC AAH78369:  
 XX 26-NOV-2001 (first entry)  
 DT  
 XX Nucleotide sequence of a human protein kinase/protein phosphatase.  
 DE Human; protein kinase; protein phosphatase; signal transduction; ss.  
 XX  
 KM Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FH 33..2630  
 FT /tag= a  
 FT /product= "protein kinase/protein phosphatase"  
 XX  
 PV MO200109316-A1.  
 XX  
 PD 08-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000MO-JP05061.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 XX 18-OCT-1999; 99US-0159590.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 17-FEB-2000; 2000US-0183322.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 XX (HELI-) HELIX RES INST.  
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX Iihil S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;  
 PI Senoo C, Nezu J;  
 XX WPI: 2001-570286/64.  
 DR P-PSDB; AAG67610.  
 XX  
 PT New genes encoding proteins with protein kinase/protein phosphatase  
 PT activity, useful in the diagnosis and treatment of diseases -  
 XX  
 PS Claim 1: Page 67-76; 233pp; Japanese.  
 XX  
 CC The present sequence encodes a human protein kinase/protein phosphatase.  
 CC It is expected that the protein kinase/protein phosphatase gene  
 CC participates in signal transduction in cells. The protein kinase/protein  
 CC phosphatase polypeptides and polynucleotides are useful for developing  
 CC diagnostics and treatment agents for human and animal diseases. The  
 CC protein kinase/protein phosphatase polypeptides are useful as target  
 CC molecules in designing novel drugs. The protein kinase/protein  
 CC phosphatase polynucleotides are useful as a source of probes and  
 CC primers, which may be used to isolate homologous sequences.  
 XX  
 SQ Sequence 2718 BP; 840 A; 586 C; 688 G; 604 T; 0 other;  
 QY  
 Query Match 43.5%; Score 2476.8; DB 22; Length 2718;  
 Best local similarity 96.0%; Pred. No. 0;  
 Matches 2603; Conservative 0; Mismatches 7; Indels 102; Gaps 2;  
 QY 374 ttgagtcacacttgcagcttgcagcaaatatgcttccatcattgctgacagcagtgatg 433  
 DB 1 ttgagtcacacttgcagcttgcagcaaatatgcttccatcattgctgacagcagtgatg 60  
 QY 434 accagctcctaatggccttcacacttgcagtgatgataatcttccctcaggaagc 493  
 DB 61 accagctcctaatggccttcacacttgcagtgatgataatcttccctcaggaagc 120  
 QY 494 cttaagaagaataatccacacgctgtgcaacaacagcgttccctcaggaacacatg 553

Db 121 ctcacaggaataaacccacacgctcgtlcaacaacagctcgtctccctcaagcagcaactcgt 180  
QY 554 cttggaagtccttaaaaaccctgcgcagcaagaagaaatgaaatacagaacacagcctcccgcg 613  
Db 181 ctggagatccctaaagccctgcgcagcaagaagaaagatgaataacatacaggaagcctccgcgg 240  
QY 614 atactagtaactggtcttaacccaattccagagagacaatcgtcgtcgtctcctagtcoca 673  
Db 241 atactagtaactggtcttaacccaattccagagagacaatcgtcgtcgtcctcgtagtcoca 300  
QY 674 gtccaccacatggtggcgcaagtgcgcgagaaacaagaacagatagaatggagagcaag 733  
Db 301 gtccaccacatggtggcgcaagtgcgcgagaaacaagaacagatagaatggagagcaag 360  
QY 734 aggaataacagtcgtctgaacacctgtctgacacatcgtcgcacacggtcttgagcatalacca 793  
Db 361 aggaataacagtcgtctgacacctgtctgacacatcgtcgcacacggtcttgagcatalacca 420  
QY 794 acctctctctggaatgaatctggttaagaagctccccaacgaatggagggcctctg 853  
Db 421 acctctctctggaatgaatctggttaagaagctccccaacgaatggagggcctctg 480  
QY 854 gaatccatgtagtgcttcagtgctgcagggcgagaaacccctggggttatatagtaaac 913  
Db 481 gaatccatgtagtgcttcagtgctgcagggcgagaaacccctggggttatatagtaaac 540  
QY 914 gattggaagaagtgctgaacacatgaaacatcttttcgtgaagaatgaatgcatg 973  
Db 541 gattggaagaagtgctgaacacatgaaacatcttttcgtgaagaatgaatgcatg 600  
QY 974 tcaagataatgaatgtagcagacctcgaataagaagaattgaaacaacacacacatgcttc 1033  
Db 601 tcaagataatgaatgtagcagacctcgaataagaagaattgaaacaacacacacatgcttc 660  
QY 1034 gccaaagccatgctgtaacacacacatcttggttccatgcttccctgcagcaataaagagc 1093  
Db 661 gccaaagccatgctgtaacacacacatcttggttccatgcttccctgcagcaataaagagc 720  
QY 1094 agtaagaacaacatacccaaaatgagaagaacaatactacatccaagccgttttaagccgt 1153  
Db 721 agtaagaacaacatacccaaaatgagaagaacaatactacatccaagccgttttaagccgt 780  
QY 1154 acagccagatatactgtaacaacagagtgctgtaacagtgcaagggctccaacggtgcagagag 1213  
Db 781 acagccagatatactgtaacaacagagtgctgtaacagtgcaagggctccaacggtgcagagag 840  
QY 1214 caccgcgactggaacacacccgcttgagcgatagactctcaactcaagactcaactcaagc 1273  
Db 841 caccgcgactggaacacacccgcttgagcgatagactctcaactcaagactcaactcaagc 900  
QY 1274 caacacccctcgggaaacaacacacatccgctcgaagctccgcaactcgaatglatatagta 1333  
Db 901 caacacccctcgggaaacaacacacatccgctcgaagctccgcaactcgaatglatatagta 960  
QY 1334 cgaactgtaagcagtggttataacacacaaaataagcagaagggcttaataaccagctta 1393  
Db 961 cgaactgtaagcagtggttataacacacaaaataagcagaagggcttaataaccagctta 1020  
QY 1394 agaaagtgtaacagaaggtttgggatacagcatcactccagaaatgtaaatagtggt 1453  
Db 1021 agaaagtgtaacagaaggtttgggatacagcatcactccagaaatgtaaatagtggt 1080  
QY 1454 cagctcccaatcctatgtgaanaaacatctctcccgggggcgccatctcagatgagcgcac 1513  
Db 1081 cagctcccaatcctatgtgaanaaacatctctcccgggggcgccatctcagatgagcgcac 1140  
QY 1514 ttaagtgcaagagcaagcaactatagagtgtaaatgtagattagtgggccaatcccaag 1573  
Db 1141 ttaagtgcaagagcaagcaactatagagtgtaaatgtagattagtgggccaatcccaag 1200  
QY 1574 aggaagtgcttccgctggtgagaagaacaaagatgaaaggaactcgtggaactctcgtct 1633

Db 1201 aggaagtgcttccgctggttgagaagcaccaagaatggaaggaactcgtggaactctcgtct 1260  
QY 1634 ttcgcaggaagaagcgccttcccccagaaggaactgaatgcaagacgaagccagatgcaaga 1693  
Db 1261 ttcgcaggaagaagcgccttcccccagaaggaactgaatgcaagacgaagccagatgcaaga 1320  
QY 1694 ttccaaaagaaacgaagaagacagaatgagatattgcttctaacctcgtatggcaccaag 1753  
Db 1321 ttccaaaagaaacgaagaagacagaatgagatattgcttctaacctcgtatggcaccaag 1380  
QY 1754 aattctcagatcttgaagcccaatgaatgaatgaagatctgcagggccttggtgacgtg 1813  
Db 1381 aattctcagatcttgaagcccaatgaatgaatgaatgaagatctgcagggccttggtgacgtg 1440  
QY 1814 tcaaaagtgaaacccggtcaaaaagaaacacacagatcttggaatccttgccaagtcocatla 1873  
Db 1441 tcaaaagtgaaacccggtcaaaaagaaacacacagatcttggaatccttgccaagtcocatla 1500  
QY 1874 ttaatggaagagcagcatcctaagaatggaaggtcttcgggtgaatgaatcaactgatacag 1933  
Db 1501 ttaatggaagagcagcatcctaagaatggaaggtcttcgggtgaatgaatcaactgatacag 1560  
QY 1934 taaatggaagaatccctggttggaagaaacaaacaaagatgcaltggaanaaccctaagaaggt 1993  
Db 1561 taaatggaagaatccctggttggaagaaacaaacaaagatgcaltggaanaaccctaagaaggt 1620  
QY 1994 ctatgctctactgaagacaaataaagaggaatgatccagcttattgttgcaagagagataa 2053  
Db 1621 ctatgctctactgaagacaaataaagaggaatgatccagcttattgttgcaagagagataa 1680  
QY 2054 gcaagtgcaaatgagctgaaagtgcaacctgtagagcccccctggaactgagctgcgccaatgaa 2113  
Db 1681 gcaagtgcaaatgagctgaaagtgcaacctgtagagcccccctggaactgagctgcgccaatgaa 1740  
QY 2114 cagcggttgatgatatagagaacgaagaattcccatctccctctacagtgagatgtaggggc 2173  
Db 1741 cagcggttgatgatatagagaacgaagaattcccatctccctctacagtgagatgtaggggc 1800  
QY 2174 tlgatgaatcgcgcgcgcgaacaaatgctgcgcctcagtaagaataagtgtagcagaat 2233  
Db 1801 tlgatgaatcgcgcgcgcgaacaaatgctgcgcctcagtaagaataagtgtagcagaat 1851  
QY 2234 accagctgtccctacagtgaaatgccccaaagatgacaatgcatatataagaagatgaca 2293  
Db 1852 accagctgtccctacagtgaaatgccccaaagatgacaatgcatatataagaagatgaca 1911  
QY 2294 ggttgccaatgcttccctccacatctctcgaacaggtccctccacagctcccatgatgagc 2353  
Db 1912 ggttgccaatgcttccctccacatctctcgaacaggtccctccacagctcccatgatgagc 1971  
QY 2354 tggggtttgtgaacggcagaatgctgtaacttgggccaaggtctgcaatcagtgatccagcg 2413  
Db 1972 tggggtttgtgaacggcagaatgctgtaacttgggccaaggtctgcaatcagtgatccagcg 2031  
QY 2414 actgctctttgagtcagaatggttgatccaggtctctgctttcaacgaagaagaattggac 2473  
Db 2032 actgctctttgagtcagaatggttgatccaggtctctgctttcaacgaagaagaattggac 2091  
QY 2474 gtccagatgctgcagaanaacgcacaagaacatcttcagatgcagatccaattggatttg 2533  
Db 2092 gtccaga----- 2097  
QY 2534 ttaaaaacagaaatacaaaaagcatgattagtgatactgacagagacataaactcaata 2593  
Db 2098 -----ctgacgcagactaaactcaata 2118  
QY 2594 cagtggaatgaacagaagaaggttccccaagcagagaatgttggtctctccctcgtggtctga 2653  
Db 2119 cagtggaatgaacagaagaaggttccccaagcagagaatgttggtctctccctcgtggtctga 2178  
QY 2654 agaaagcagcctcgttggaagagctcgcagaccgaatcttcgaagtgagacttgaaatgggg 2713  
Db 2179 agaaagcagcctcgttggaagagctcgcagaccgaatcttcgaagtgagacttgaaatgggg 2238

QY	2714	atatctctccatcgttcacgccgacgagataatcagagcgagggtgtgcaatgagagct	2773
Db	2239	atatctctctccatcgttcacgccgacgagataatcagagcgagggtgtgcaatgagagct	2298
QY	2774	tcagagctgcgcatacgcagaaatcttatgataaaccccggttagatgatgatgataagca	2833
Db	2299	tcagagctgcgcatacgcagaaatcttatgataaaccccggttagatgatgatgataagca	2358
QY	2834	tggagaccttggagaagagacacagaagaagctcaagatlcagggagagagctgtatcca	2893
Db	2359	tggagaccttggagaagagacacagaagaagctcaagatlcagggagagagctgtatcca	2418
QY	2894	cagcgagtgatcagccttccacactctctcggagagacaaatgaaatgagaaacgaagaag	2953
Db	2419	cagcgagtgatcagccttccacactctcggagagacaaatgaaatgagaaacgaagaag	2478
QY	2954	gtgataagactgtatagaataaagaagataaactcgtgaagaagaagaagaagatagagata	3013
Db	2479	gtgataagactgtatagaataaagaagataaactcgtgaagaagaagaagaagatagagata	2538
QY	3014	agggagaagataaagaaagcaagaagggagatctgtgaaggcttgggaagactgttcca	3073
Db	2539	agggagaagataaagaaagcaagaagggagatctgtgaaggcttgggaagactgttcca	2598
QY	3074	ggtttggcaaac 3085	
Db	2599	gaccttggcaaac 2610	

Result	4
AD01182	
ID	AA001182 standard; cDNA; 2554 BP.
XX	
AC	AA01182;
XX	
DT	04-OCT-2000 (first entry)
XX	
DE	PDZ domain-comprising PHIP (ephrin Interacting Protein) complex cDNA.
XX	
KW	PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
KW	cellular process; nerve cell interaction; regeneration of nerve cell;
KW	axonogenesis; antiproliferative; proliferative disorder; treatment; CRIP;
KW	differentiative disorder; PHIP; ephrin interacting protein; synenin; ss
XX	
OS	Mus sp.
XX	
FH	Location/Qualifiers
FT	Key
FT	306..2540
FT	CD5
FT	/*tag= a
FT	/product= "PHIP (ephrin Interacting Protein) complex"
FT	/note= "Homologous to C. elegans PAR-3 protein"
XX	
XX	
PN	WO200031124-A2.
XX	
PD	02-JUN-2000.
XX	
PE	19-NOV-1999; 99WO-CA01101.
XX	
XX	20-NOV-1998; 98US-0109158.
PR	(MOUN ) MOUNT SINAI HOSPITAL.
PA	
XX	
XX	
PI	Lin D, Pawson A;
DR	
XX	WPI; 2000-400038/34.
DR	P-PSDB; AA171410.
XX	
PT	Isolated complex for treating proliferative or differentiative
XX	disorders comprises B class ephrin and PDZ domain containing protein -
PS	disclosure; Page 57-59; 59pp; English.
XX	

The patent discloses a complex comprising of a B class ephrin and PDZ domain containing protein. B class ephrins function as ligands for Eph receptor tyrosine kinases (RTK) and possess a transmembrane element and a highly conserved hydrophobic tail at the C-terminus that contains a PDZ binding site. This complex is used in methods to modulate the interaction of a B class ephrin and PDZ domain containing protein and to identify modulator compounds. It is also used for modulating cellular processes like, axonogenesis, nerve cell interactions and regeneration or cells. The complex is also useful for treating proliferative or differentiative disorders associated with this protein complex. The present sequence is the cDNA clone encoding Phip (Ephrin interacting Protein) complex, comprising three PDZ domains. This clone is isolated by screening a lambdaclonex Caenorhabditis elegans PAR-3 protein involved Phip is closely related to the early embryo. Phip cDNA fragment was used in regulating polarity of the early embryo. Phip cDNA fragment proteins as a probe, to isolate other transcripts like Grip and synenin proteins from 10.5 day mouse embryo library.

Sequence 2554 BP; 650 A; 697 C; 745 G; 462 T; 0 other;

Query Match	29.8%;	Score 1695.8;	DB 21;	Length 2554;
Best Local Similarity	84.8%;	Pred. No. 0;	Mismatches 342;	Indels 0; Gaps 0;
Matches 1901;	Conservative	0;		
QY 1	atgaagaatgcacgctgctgcttcggagagaccgcggcgctgcctgcctgcctgcgggagagccac	60		
Db 306	atgaagaatgcacgctgcttcggagagaccgcggcgctgcctgcctgcctgcgggagagccgc	365		
QY 61	atgaagaatgcacacccctccatccagcagcgctgtaaccgcctaccggaaaggacatccgaag	120		
Db 366	atgaagaatgcacacccctccatccagcagcgctgtaaccgcctaccggaaaggacatccgaag	425		
QY 121	gatccaactactgagatcacagctgctacgcctctggaacatggaatggaagaaactactaac	180		
Db 426	gatccaactactgagatcacagctgctacgcctctggaacatggaatggaagaaactactaac	485		
QY 181	cttgatgacatctctctgtgtagctgagcaagaataaagacagactgtagcagtgtttgt	240		
Db 486	cttgatgacatctctctctgtgtagctgagcaagaataaagacagactgtagcagtatttgal	545		
QY 241	gagcagatcatcacatcacgagagctgtagtcacacagctgacatttcacagggatcacagac	300		
Db 546	gaaacagatcatccccaatcagagagagctgtgtacccaagcagacgtccacacggagaaaccagat	605		
QY 301	ccaagagatatttgtagtagagctctggaaccaaccaatgctcagccttcacaccttaccac	360		
Db 606	ccaagagatatttcggacagctgagctctggaaccaaccaatgctcagccttcacaccttaccac	665		
QY 361	ggaacaatgtaaatgtaggtlcaacaacttcagcttccttcggacgaataatgctcttcacat	420		
Db 666	ggaacaatgtaaatgtaggtlcaacaacttcagcttccttcggacgaataatgctcttcacat	725		
QY 421	cgaacgactgtagtgagacacagctcaaatctgctccacactcttcagtgtagtaattt	480		
Db 726	cgcgcggagacagcagacccacgcttaactctgctcttcacactcttcagtgtagtaactt	785		
QY 481	tccctctgaagagccttcaagaagaaatcccaacacgctggtlcaacaacagctggcttcctc	540		
Db 786	tccctctgaagagccttcaagaagaaatcccaacacgctggtlcaacaacagctggcttcctc	845		
QY 541	aagcagaacacgctcgtggaagctcctaaccctctgcgcgaagaagaagatgaaactacaga	600		
Db 846	aagcagaacacacgctcgtggaagctcctaaccctctgcgcgaagaagaagatgaaactacaga	905		
QY 601	agcctccgcgggtactagtaactgcttaaccaattccagagagcaaatgctgcctcg	660		
Db 906	agcctccacagggaccccaatgactgtgttccaaacagttccagagagacaacgcgcctcc	965		
QY 661	tctctagatgcacgtaccaccaatgctggtggcaagtgtgtgagaaataaagacagatgag	720		
Db 966	tccctctagatgcacgtaccaccaatgctgagacggtgtgccttgtagaagaacaagacagatgag	1021		

QY 721 gatggagacagaagagataacaagtcgtgttaaacgtgttgacatgctgaacagggttg 780  
DB 1026 gaaggacagaagaagacagcagccgagttggaccggttggacatgctgacatcgatg 1085  
QY 781 gagacataccaacttctctctgataatgtaaaagcgtgtgaagctcccaagat 840  
DB 1086 gagaacatgcccacaacttctccctcgaatgataatgtaaaagcgtgtgaagctcccaagat 1145  
QY 841 ggaaggcctctggaatccatgtaatgctctcaatgctcgaagcgaggaacccctggg 900  
DB 1146 ggaaggccctgggaatccatgtaatgctctcaatgctcgaagcgaggaacccctggg 1205  
QY 901 ttatagtaaaaacgacttggaagaagtggttaaaagcgtgaacatgtaaaatcttctgag 960  
DB 1206 ttgttagtaaaagcggttggaaagaagcggttaagcgttggaagaacatcttccatgag 1265  
QY 961 aatgatgcatgtgtaagataatgataatgagcagccttgaaatagaagaattggaacaaga 1020  
DB 1266 aatgactgcatgtggaagataatgaacgaatggaatcttgcgaacaagaatltgagaacga 1325  
QY 1021 caacatagttccgaagcagatgctgaacccatcatcttgctcatgtgtcttga 1080  
DB 1326 caacatagttccgaagcagatgctgagtgctgagtgctcatgtgtcttga 1385  
QY 1081 gcaaatlaaagcagatgaacaaactlcccaagaatggaagaacaactactatcaagc 1140  
DB 1386 gcaaatlaaagcagatgaacaaactlcccaagaatggaagaacaactactatcaagc 1445  
QY 1141 gcttttagccttgacagcagatataatgtaaaagcagatggaagaagcagcggttgaac 1200  
DB 1446 gcttttagccttgacagcagatataatgtaaaagcagatggaagaagcagcggttgaac 1505  
QY 1201 acggtgcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1260  
DB 1506 gcatggccagagacacccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1565  
QY 1261 ctactctatgacgaacccctcggaagaacacacacacacacacacacacacacacacac 1320  
DB 1566 ctactctatgacgaacccctcggaagaacacacacacacacacacacacacacacacac 1625  
QY 1321 aatgtaattagtaagcagtgtaagcagtggttataacacaaataatggaagaagcgtt 1380  
DB 1626 agtgtgctatgatacacaagtgacagtggttataacacaaataatggaagaagcgtt 1685  
QY 1381 aatatccagcttaagaaggtacagaaggttggatccagcatcactccagagatgta 1440  
DB 1686 aatatccagcttaagaaggtacagaaggttggatccagcatcactccagagatgta 1745  
QY 1441 acaataggtgctcagcctcccaatctatgtgaataaacttcccccggggggcgagcatt 1500  
DB 1746 acaataggtgctcagcctcccaatctatgtgaataaacttcccccggggggcgagcatt 1805  
QY 1501 cagatgcccagacttaagaagcagagacttatagaggttaaatggaggtatgtagt 1560  
DB 1806 cagatgcccagacttaagaagcagagacttatagaggttaaatggaggtatgtagt 1865  
QY 1561 ggcgaatcccaagaaggtgttctgctgtgtgaagaacacagaatggaagaagcgtg 1620  
DB 1866 ggcgaatcccaagaaggtgttctgctgtgtgaagaacacagaatggaagaagcgtg 1925  
QY 1621 agccttctggtcttgcgaagaagaagccttccacccaagggaactgaatgacagcca 1680  
DB 1926 agccttctggtcttgcgaagaagaagccttccacccaagggaactgaatgacagcca 1985  
QY 1681 agcgaatgcatcctcaaaagaacgaagaagcagagatggaagatggttcttaccct 1740  
DB 1986 agcgaatgcatcctcaaaagaacgaagaagcagagatggaagatggttcttaccct 2045  
QY 1741 gatggcagcaggaattctgacatttgaatltcccaatgaatgacatgacatgctcagcg 1800  
DB 2046 gatggcagcaggaattctgacatttgaatltcccaatgacatgacatgctcagcg 2105  
QY 1801 ctggtgtcagtgtaacgaagtaacccgttcaaaagaacacacagatltgggaattctt 1860

DB 2106 ctggtgtcagtgtaacgaagtggaacggttccaaagaagaacacagatltgggacatcc 2165  
QY 1861 gtcgaatgcatcttaattgaatggaagcagacatctaaagaatggaagcgttggatgatt 1920  
DB 2166 gttaaatccatcatcaatggtggaagcgtgcatctaaagaatggaagcgttggatgatt 2225  
QY 1921 caactgataagcagtaaatggaatccctgtgtgggaagaacaaagaatgcatgga 1980  
DB 2226 cagctgataagcgtgtgaatggaatctctactgtggcaagaacaaagaatgcatgga 2285  
QY 1981 accctaagaaggtctatgtctactgaaggcaataaagaagaatgcatgattgtt 2040  
DB 2286 actctaggaaggttccatgtccacccgaagaacagcgtgcatgcatcagatgtg 2345  
QY 2041 gcaaggagaataagcagagcagatgagcgtgaagctgaacccctggagccctgagctgag 2100  
DB 2346 gcaaggcagatcagagatgtaacgagcgtggttccctggagccctgagctgag 2405  
QY 2101 ctgcccattgaaacagcgttggatgataagagaacgaagaatltcccatcctcagat 2160  
DB 2406 ctgcccattgaaacagcgttggatgataagagaacgaagaatltcccatcctcagat 2465  
QY 2161 gggatgagggcgtgataatcgcacagagaatgctgcccctcagtagtaatgggt 2220  
DB 2466 gggatgagggcgtgataatcgcacagagaatgctgcccctcagtagtaatgggt 2525  
QY 2221 ggtgaggttaatacagcgttc 2243  
DB 2526 ggtgaggttaatacagcgttc 2548

RESULT 5  
AAH72793  
ID AAH72793 standard; cDNA; 1800 BP.  
XX  
AC AAH72793;  
XX  
DT 19-SEP-2001 (first entry)  
XX  
DE Human cervical cancer marker nucleic acid 4067.  
XX  
KM Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200142467-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 08-DEC-2000; 2000WO-US33312.  
XX  
PR 08-DEC-1999; 99US-0169681.  
PR 21-DEC-1999; 99US-0171350.  
PR 14-MAR-2000; 2000US-0189315.  
PR 12-MAY-2000; 2000US-0203791.  
PR 09-JUN-2000; 2000US-0210600.  
PR 21-JUN-2000; 2000US-0220114.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Deeds J, Berger A, Zhao X;  
XX  
DR WPI; 2001-375006/39.  
XX  
PR New isolated nucleic acid for diagnosing and treating cervical cancer  
PR and for assessing and detecting compounds for treating the cancer -  
XX  
PS Claim 1; Page 855-856; 1051p; English.  
XX  
CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with



CC cervical cancer or has a pre-malignant condition, to monitor the  
 CC progression of cervical cancer or a premalignant condition in a patient;  
 CC and to select and/or assess the efficacy of a compound or therapy for  
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
 CC useful for gene therapy.  
 XX

Sequence 1800 BP; 578 A; 359 C; 466 G; 394 T; 3 other;

Query Match Best Local Similarity 28.8%; Score 1638.6; DB 22; Length 1800;  
 Matches 1660; Conservative 0; Mismatches 4; Indels 9; Gaps 1;

QY 1413 gggatcagcactccctccagagatgttaacaataagtggtcagctccatctatgtgaa 1472  
 Db 3 gggatcagcactccctccagagatgttaacaataagtggtcagctccatctatgtgaa 1472  
 QY 1473 aaacattctcccccggggggggccatcagagatggccagacttaaggcagagagagact 1532  
 Db 63 aaacattctcccccggggggggccatcagagatggccagacttaaggcagagagagact 1532  
 QY 1533 tatagagtaaatggagtagatattggtggcaatcccaagaggaagtgtgttcgtgtt 1592  
 Db 123 tatagaggttaaatggagtagatattggtggcaatcccaagaggaagtgtgttcgtgtt 1592  
 QY 1593 gagaagccaccagaatggaagaaactgtgagccttcgtgttccttcgccaagagagcgtt 1652  
 Db 183 gagaagccaccagaatggaagaaactgtgagccttcgtgttccttcgccaagagagcgtt 1652  
 QY 1653 ccaccacaagggaactgcatgacagcccaagcagatgcagattcccaaaaagacgaaagc 1712  
 Db 243 ccaccacaagggaactgcatgacagcccaagcagatgcagattcccaaaaagacgaaagc 1712  
 QY 1713 aagaagatagatatgttcttaacactgtgtgcaacagaggaattcttacaatttgaagt 1772  
 Db 303 aagaagatagatatgttcttaacactgtgtgcaacagaggaattcttacaatttgaagt 1772  
 QY 1773 cccacttaattgattcagagatctgcagagccttggtgtcagtgtcacaaggttaacgcgttcaa 1832  
 Db 363 cccacttaattgattcagagatctgcagagccttggtgtcagtgtcacaaggttaacgcgttcaa 1832  
 QY 1833 agaaagacagcagagatttgggaactcttctgcaagtcacattatattatggagagacagc 1892  
 Db 423 agaaagacagcagagatttgggaactcttctgcaagtcacattatattatggagagacagc 1892  
 QY 1893 taaagatggaaggcttcgggtgaatgataactgtatagcagtaaatggaagatccctgtt 1952  
 Db 483 taaagatggaaggcttcgggtgaatgataactgtatagcagtaaatggaagatccctgtt 1952  
 QY 1953 gggcagaagacaacccaagatgccaatggaaccctaagaaggtctatgtctcttaaggca 2012  
 Db 543 gggcagaagacaacccaagatgccaatggaaccctaagaaggtctatgtctcttaaggca 2012  
 QY 2013 taaacgaaggaatgatacagcttattgttgcagaaggaataagcaagtgcatagtcga 2072  
 Db 603 taaacgaaggaatgatacagcttattgttgcagaaggaataagcaagtgcatagtcga 2072  
 QY 2073 gtccacttggagccccccttgagccttgagctgcccattgaaacagcgcttgcagcagag 2132  
 Db 663 gtccacttggagccccccttgagccttgagctgcccattgaaacagcgcttgcagcagag 2132  
 QY 2133 acgaagaattcccatcctctacagtgaggttggagggccttgatgtaacgcagccagag 2192  
 Db 723 acgaagaattcccatcctctacagtgaggttggagggccttgatgtaacgcagccagag 2192  
 QY 2193 aaatgctgcctcagtagagataatggatgcaggttaataaccgcgcttccccaagct 2252  
 Db 783 aaatgctgcctcagtagagataatggatgcaggttaataaccgcgcttccccaagct 2252  
 QY 2253 gaatatgcccccaagatgacactgtcatatagaagaataacaggttgcagtgctctcc 2312  
 Db 834 gaatatgcccccaagatgacactgtcatatagaagaataacaggttgcagtgctctcc 2312

QY 2313 acatctctgacacagtcctctccagctccacatgatatgtggttctgacggcaga 2372  
 Db 894 acatctctgacacagtcctctccagctccacatgatatgtggttctgacggcaga 2372  
 QY 2373 tgcctgacttgggccaaggtctgcaatcagtgattcaacgagcgtcttggatccaga 2432  
 Db 954 tgcctgacttgggccaaggtctgcaatcagtgattcaacgagcgtcttggatccaga 2432  
 QY 2433 tgtttaccagttctgtcttccacaggaaggtttggacgtcagaatgtcagaana 2492  
 Db 1014 tgtttaccagttctgtcttccacaggaaggtttggacgtcagaatgtcagaana 2492  
 QY 2493 acgcacaagcagcatttcaagatgcacgtcaatttgatctgtttaaaccgaaatcaaa 2552  
 Db 1074 acgcacaagcagcatttcaagatgcacgtcaatttgatctgtttaaaccgaaatcaaa 2552  
 QY 2553 aagcatgatttagtatagtcagcagactaaactcaatcaagtgatgacagagaagc 2612  
 Db 1134 aagcatgatttagtatagtcagcagactaaactcaatcaagtgatgacagagaagc 2612  
 QY 2613 agttctcccgacagagaatgtggtccttcctcctggttcgaaagatgaagctgttga 2672  
 Db 1194 agttctcccgacagagaatgtggtccttcctcctggttcgaaagatgaagctgttga 2672  
 QY 2673 gactctgcagacgcagcttgcagaggtgacttggatgggatattccttccatctcc 2732  
 Db 1254 gactctgcagacgcagcttgcagaggtgacttggatgggatattccttccatctcc 2732  
 QY 2733 acgcccgcgagataatagaaggcagggagatgcaatgagcttcagagctcagtcagca 2792  
 Db 1314 acgcccgcgagataatagaaggcagggagatgcaatgagcttcagagctcagtcagca 2792  
 QY 2793 attctatgataaccgcggtgtagatgcatgcatggaagcattggaagaaaga 2852  
 Db 1374 attctatgataaccgcggtgtagatgcatgcatggaagcattggaagaaaga 2852  
 QY 2853 cacagaagaagaattcaagaatacagggagagagctgtatccacagcagtgatcagcttc 2912  
 Db 1434 cacagaagaagaattcaagaatacagggagagagctgtatccacagcagtgatcagcttc 2912  
 QY 2913 cactctctggagagacaatgaaatgaaacgaaggaagtgatgaagaactgtatagaa 2972  
 Db 1494 cactctctggagagacaatgaaatgaaacgaaggaagtgatgaagaactgtatagaa 2972  
 QY 2973 aaagataaaacttgaagaagaagaagaatagatagaataaggagaaggtataaagaa 3032  
 Db 1554 aaagataaaacttgaagaagaagaagaatagatagaataaggagaaggtataaagaa 3032  
 QY 3033 agccaagaaggaagctggaaggtcttggagacatgttccaggttggcaaac 3085  
 Db 1614 agccaagaaggaagctggaaggtcttggagacatgttccaggttggcaaac 3085  
 RESULT 6  
 AAF91875 standard: cDNA; 867 BP.  
 AC AAF91875;  
 XX  
 DT 22-MAY-2001 (first entry)  
 DE Human secreted protein-encoding gene 18 cDNA clone HFPK57, SEQ ID NO:28.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW cardiovascular disorder; psoriasis; sepsis; diabetes; kidney disorder;  
 KW gastrointestinal disorder; angiogenic disorder; kidney disorder;  
 KW endocrine disorder; infection; pregnancy-related disorder;  
 KW cell culture; chemotaxis; food additive;



binding partner identification: ss.

XX Homo sapiens.

XX WO200118022-A1.

XX 15-MAR-2001.

XX 31-AUG-2000; 2000WO-US24008.

XX 03-SEP-1999; 99US-0152315.

XX 03-SEP-1999; 99US-0152317.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ni J, Baker KP, Birse CE, Fisceella M, Komatsoulis GA, Rosen CA;  
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Latleur DW;  
PI Moore PA, Shi Y, Wei Y, Florence KA;

XX WPI: 2001-203081/20.  
XX P-PSDB: AAB87359.

XX Nucleic acid molecules encoding human secreted proteins, used in  
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
PT Parkinson's diseases and cancers -

XX Claim 1: Page 498; 607pp; English.

XX AAF91588-AAF91929 represent cDNAs corresponding to 52 human secreted  
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.  
CC AAB87414-AAB87454 represent human secreted protein fragments. The genes  
CC and their corresponding secreted proteins are useful for preventing,  
CC treating or ameliorating medical conditions, e.g., by protein or gene  
CC therapy. Pathological conditions can be diagnosed by determining the  
CC amount of the new protein in a sample or by determining the presence of  
CC mutations in the new genes. Specific uses are described for each of the  
CC 52 genes, based on the tissues in which they are most highly expressed,  
CC and include developing products for the diagnosis or treatment of  
CC proliferative disorders, cancer, tumours, foetal and developmental  
CC abnormalities, haematopoietic disorders, diseases of the immune system,  
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
CC allergies, neurological disorders (e.g., Alzheimer's disease,  
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein-encoding cDNA of the invention.

XX SQ Sequence 867 BP; 233 A; 180 C; 197 G; 256 T; 1 other;

Query Match 14.8%; Score 842; DB 22; Length 867;  
Best Local Similarity 100.0%; Pred. No. 1.2e-221;  
Matches 842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4848 ctgcacatggccctgtctatcatcagagggagagctaaagaagaattctctaggaag 4907  
DB 13 ctgcacatggccctgtctatcatcagagggagagctaaagaagaattctctaggaag 72  
QY 4908 agctaatggccacagatcactagatatttaattagtttgcctgcacagcttgt 4967  
DB 73 agctaatggccacagatcactagatatttaattagtttgcctgcacagcttgt 132

QY 4968 caggaaaggcacagaatgagacagataaaccagaacagctatcttgatctctctac 5027  
DB 133 caggaaaggcacagaatgagacagataaaccagaacagctatcttgatctctctac 192  
QY 5028 ggttttcaagtcagaggaattgtctgtctaatgtatgtatccaaactgtcatgtcga 5087  
DB 193 ggttttcaagtcagaggaattgtctgtctaatgtatgtatccaaactgtcatgtcga 252  
QY 5088 ctggcgatgccacgctccctaaagtaattctgcacatgaataaagaagaagaagcc 5147  
DB 253 ctggcgatgccacgctccctaaagtaattctgcacatgaataaagaagaagaagcc 312  
QY 5148 gttaacacatcacacagagaacatttcgggtcccaacagcgtgtgtgcaggaagctact 5207  
DB 313 gttaacacatcacacagagaacatttcgggtcccaacagcgtgtgtgcaggaagctact 372  
QY 5208 ctccgctcagatattgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5267  
DB 373 ctccgctcagatattgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 432  
QY 5268 ttcatgttctgaacatcctctgtatgttaacacatgctgggtgtgtgtgtgtgtgtgtga 5327  
DB 433 ttcatgttctgaacatcctctgtatgttaacacatgctgggtgtgtgtgtgtgtgtgtga 492  
QY 5328 tcccgatgtgaaanaagcgtgaggtgaagctcagacatccatgtattactttaanaac 5387  
DB 493 tcccgatgtgaaanaagcgtgaggtgaagctcagacatccatgtattactttaanaac 552  
QY 5388 aaaaaaagaacatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 5447  
DB 553 aaaaaaagaacatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 612  
QY 5448 gttagctgttttgaagaatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 5507  
DB 613 gttagctgttttgaagaatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 672  
QY 5508 gtgtcaatgtctctgttcttaatacagtcggtggtgtgtgtgtgtgtgtgtgtgtgtac 5567  
DB 673 gtgtcaatgtctctgttcttaatacagtcggtggtgtgtgtgtgtgtgtgtgtgtgtac 732  
QY 5568 ggcgagtgacctgtgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 5627  
DB 733 ggcgagtgacctgtgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 792  
QY 5628 agcttctctgtatcatatgtcaactcctaataaaggacatattcttcggttaaaaaa 5687  
DB 793 agcttctctgtatcatatgtcaactcctaataaaggacatattcttcggttaaaaaa 852  
QY 5688 aa 5689  
DB 853 aa 854

RESULT 7  
AAS27320  
ID AAS27320 standard; cDNA: 823 BP.  
XX AAS27320;  
XX 07-NOV-2001 (first entry)  
XX cDNA encoding novel signal transduction pathway protein, Seq ID 355.  
DE Neuroprotective; cytosolic; dermatological; immunosuppressive; tumour;  
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;  
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;  
KW acquired immune deficiency syndrome.

XX Homo sapiens.  
OS  
XX  
PN WO200154733-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01312.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216680.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 18-AUG-2000; 2000US-0225759.  
PR 22-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227099.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249220.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0259678.  
XX

PA (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR MPI: 2001-465460/50.  
 DR P-PSDB: AAU17403.  
 XX  
 PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
 PR promoting disorders related to the proteins, including cancers, immune  
 PT disorders and neuronal disorders  
 XX  
 PS Claim 1: SEQ ID No 355; 880pp; English.

XX  
 CC The invention relates to novel isolated polypeptides (I), and  
 CC polynucleotides (II). (I), (II) are useful for  
 CC diagnosing, preventing and treating diseases including immune system  
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
 CC transplant rejections and graft versus host disease, infectious diseases  
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
 CC other blood-related disorders (sickle cell anaemia), myeloproliferative  
 CC disorders, primary haematopoietic disorders, hyperproliferative  
 CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
 CC disorders (e.g. Alzheimer's disease and Parkinson's disease), chromosomal  
 CC abnormalities (Down syndrome), ischemic injury (e.g. stroke), renal  
 CC disorders (e.g. glomerulonephritis), cardiovascular disorders  
 CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in  
 CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
 CC Addison's disease), reproductive system disorders, gastrointestinal  
 CC disorder (inflammatory disorders), liver disorders (cirrhosis),  
 CC as stimulators of B-cell responsiveness to pathogens, activators of  
 CC T-cells, to induce higher affinity antibodies, and as a means to induce  
 CC tumour proliferation in pathologies e.g. acquired immune deficiency  
 CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction  
 CC pathway protein coding sequences and PCR primers of the invention.  
 XX

Query Match 14.4%; Score 819.4; DB 22: Length 823;  
 Best Local Similarity 99.6%; Pred. No. 2e-215;  
 Matches 820; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 773 cgggtttgagagatatacccaactttctctgatatgataatgtaagctcgtagaatc 832  
 DB 1 cggntttgagagatatacccaactttctctgatatgataatgtaagctcgtagaatc 60  
 QY 833 ccaacgatgagggcctcgtggaatccatgtagtcttcaatgctcgtcagcgagaa 892  
 DB 61 ccaacgatgagggcctcgtggaatccatgtagtcttcaatgctcgtcagcgagaa 120  
 QY 893 ccctggggttattagtaaaacgatitggagaaggtgtgtaaacgtgaacatgaaatc 952  
 DB 121 ccctggggttattagtaaaacgatitggagaaggtgtgtaaacgtgaacatc 180  
 QY 953 ttccgtgagaatgattgcatgtcgaagataatgataagtcgaactcgtgaataagaattg 1012  
 DB 181 ttccgtgagaatgattgcatgtcgaagataatgataagtcgaactcgtgaataagaattg 240  
 QY 1013 aacaacacacacatgatttgcgcaagcatggtacaccatcatttggttccatgtg 1072  
 DB 241 aacaacacacacatgatttgcgcaagcatggtacaccatcatttggttccatgtg 300  
 QY 1073 ttccgtgagaatgattgcatgtcgaagataatgataagtcgaactcgtgaataagaattg 1132  
 DB 301 ttccgtgagaatgattgcatgtcgaagataatgataagtcgaactcgtgaataagaattg 360  
 QY 1133 attcaagcggttttagccttgacagccagatatttgacaacagagatgtgaaacgtgca 1192  
 DB 361 attcaagcggttttagccttgacagccagatatttgacaacagagatgtgaaacgtgca 420  
 QY 1193 ggttcaacacgtgtgagagagaccccgagctgaacaccccgctgtggaataagactc 1252  
 DB 421 ggttcaacacgtgtgagagagaccccgagctgaacaccccgctgtggaataagactc 480

QY 1253 actcaagactactcatagcgcacacccctcgggaaacacacatccgctccagctcgg 1312  
 DB 481 actcaagactactcatagcgcacacccctcgggaaacacacatccgctccagctcgg 540  
 QY 1313 cactcaagaatgattattagtagcactgtgaacgctgtgttaatacaccacaaataagga 1372  
 DB 541 cactcaagaatgattattagtagcactgtgaacgctgtgttaatacaccacaaataagga 600  
 QY 1373 agaggttaataatccagcttaagaagaaggtacagaaggttgggattcaagataactcca 1432  
 DB 601 agaggttaataatccagcttaagaagaaggtacagaaggttgggattcaagataactcca 660  
 QY 1433 gagaatgtaacaataggttgctcagctcccaatctatgtgaaacacatctcccccggggg 1492  
 DB 661 gagaatgtaacaataggttgctcagctcccaatctatgtgaaacacatctcccccggggg 720  
 QY 1493 cggccattcaaggttgccgagcttaagcagagacagactatagaggttaagtgaagtag 1552  
 DB 721 cggccattcaaggttgccgagcttaagcagagacagactatagaggttaagtgaagtag 780  
 QY 1553 attagtgagcaaatcccaagaggaagtgttgcgtgtgag 1595  
 DB 781 gtttagtggcaaatcccaagaggaagtgttgcgtgtgag 823

RESULT 8  
 AAH33594  
 ID AAH33594 standard; cDNA; 823 BP.  
 XX  
 AC AAH33594;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:650.  
 XX  
 KW Human: colon cancer; colon cancer antigen; diagnosis: detection;  
 KW colorectal carcinoma; ss.  
 XX  
 KW Homo sapiens.  
 XX  
 PN MO200122920-A2.  
 XX  
 PD 05-APR-2001.  
 XX  
 PF 28-SEP-2000; 2000MO-US26524.  
 XX  
 PR 29-SEP-1999; 99US-0157137.  
 PR 03-NOV-1999; 99US-0163280.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
 XX  
 DR MPI: 2001-235357/24.  
 DR P-PSDB: AAG74163.  
 XX  
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 PT useful for preventing, diagnosing and/or treating colorectal cancers -  
 PT  
 PS Claim 1: Page 2675; 9803pp; English.  
 XX  
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patients own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated Ps,  
 CC by inserting the nucleic acids into a host cell and culturing the cell



QY 434 acccagcttaattggccttcacactctctgtaagtagtaatttctcctgaagagc 493  
|||||  
DB 61 acccagcttaattggccttcacactctctgtaagtagtaatttctcctgaagagc 120  
QY 494 cttcaagaaaaatcccaacgcgtgtcaacaacagctgtcctccaagcagaacacgt 553  
|||||  
DB 121 cttcaagaaaaatcccaacgcgtgtcaacaacagctgtcctccaagcagaacacgt 180  
QY 554 ctggaggtccttaaaacctgcagacagaagaatgaaaaactacagaagcctccgcgg 613  
|||||  
DB 181 ctggaggtccttaaaacctgcagacagaagaatgaaaaactacagaagcctccgcgg 240  
QY 614 atactagtaactggttaaccaatttcagagagacaatgtcgtctctgaatgccca 673  
|||||  
DB 241 atactagtaactggttaaccaatttcagagagacaatgtcgtctctgaatgccca 300  
QY 674 gtcaaccaatgtgtggcaagctgtgcagaaagaagaacagatggatggagcagaag 733  
|||||  
DB 301 gtcaaccaatgtgtggcaagctgtgcagaaagaagaacagatggatggagcagaag 360  
QY 734 aggataacagctcgtgtgaacctgtgtgacatgtctgacacggtttcgaagcatalacca 793  
|||||  
DB 361 aggataacagctcgtgtgaacctgtgtgacatgtctgacacggtttcgaagcatalacca 420  
QY 794 actttctctgtagatgataatgtaaaagctcgttagaagtcctcccaacgataggagcctctg 853  
|||||  
DB 421 actttctctgtagatgataatgtaaaagctcgttagaagtcctcccaacgataggagcctctg 480  
QY 854 gaatccatgtagtccttcagtgctcgaagcgagcagaacctggggttaattagtaaac 913  
|||||  
DB 481 gaatccatgtagtccttcagtgctcgaagcgagcagaacctggggttaattagtaaac 540  
QY 914 gattggaagaaggtgg 929  
|||||  
DB 541 gattggaagaaggtgg 556

RESULT 10  
AAH71480  
ID AAH71480 standard; cDNA; 597 BP.  
XX  
AC AAH71480;  
XX  
DT 19-SEP-2001 (first entry)  
XX  
DE Human cervical cancer marker nucleic acid 2754.  
XX  
KM Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200142467-A2.  
XX  
PD 14-JUN-2001.  
XX  
PE 08-DEC-2000; 2000MO-US33312.  
XX  
PR 08-DEC-1999; 99US-0169681.  
XX  
PR 21-DEC-1999; 99US-0171350.  
XX  
PR 14-MAR-2000; 2000US-0189315.  
XX  
PR 12-MAY-2000; 2000US-0203791.  
XX  
PR 09-JUN-2000; 2000US-0210600.  
XX  
PR 21-JUL-2000; 2000US-0220114.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Deeds J, Berger A, Zhao X;  
XX  
DR WPI; 2001-375006/39.  
XX  
PT New isolated nucleic acid for diagnosing and treating cervical cancer  
XX and for assessing and detecting compounds for treating the cancer -

PS - Claim 1: Page 566; 1051pp; English.  
XX  
XX The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a pre-malignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
XX useful for gene therapy.

SO Sequence 597 BP: 180 A; 123 C; 156 G; 138 T; 0 other:

Query Match 9.1%; Score 516.2; DB 22; Length 597;  
Best Local Similarity 91.0%; Pred. No. 8.6e-132;  
Matches 576; Conservative 0; Mismatches 18; Indels 39; Gaps 1;

QY 1402 acagaaggtttggatccagcatcacttccagagatgtaacaatagttgctcagctcca 1461  
|||||  
DB 1 acagaaggtttggatccagcatcacttccagagatgtaacaatagttgctcagctcca 60  
QY 1462 atctatgtgaaaaaacattcccccggggggggccattcaagatggcgaacttaagca 1521  
|||||  
DB 61 atctatgtgaaaaaacattcccccggggggggccattcaagatggcgaacttaagca 120  
QY 1522 ggaacagacttataagatggaatggaatggaatggaatggaatggaatggaatggaat 1581  
|||||  
DB 121 ggaacagacttataagatggaatggaatggaatggaatggaatggaatggaatggaat 180  
QY 1582 gtttcgctgttggaagacccaagaatggaagaaactgtgacgttctgtcttcgcag 1641  
|||||  
DB 181 gtttcgctgttggaagacccaagaatggaagaaactgtgacgttctgtcttcgcag 240  
QY 1642 gaagacgcttccacccaagggaactgaatgcagagccaagccgaatgacgaattccaaa 1701  
|||||  
DB 241 gaagacgcttccacccaagggaactgaatgcagagccaagccgaatgacgaattccaaa 266  
QY 1702 gaaacgaagcagaagaatggaatggaatggaatggaatggaatggaatggaatggaat 1761  
|||||  
DB 267 -----gaaacgaagcagaagaatggaatggaatggaatggaatggaatggaatggaat 321  
QY 1762 acattgaagtcctcaacttaatgaltcagatcgtcagagccttgggtcagtgtaaaagt 1821  
|||||  
DB 322 acattgaagtcctcaacttaatgaltcagatcgtcagagccttgggtcagtgtaaaagt 381  
QY 1822 aacggtcacaagaagaccccgagatttgggaattcttgaagtcatattataatgga 1881  
|||||  
DB 382 aacggtcacaagaagaccccgagatttgggaattcttgaagtcatattataatgga 441  
QY 1882 ggaacagcatcacaagaatggaagcctcgggtggaatgatacactgatacagtaaatgga 1941  
|||||  
DB 442 ggggcagcatcacaagaatggaagcctcgggtggaatgatacactgatacagtaaatgga 501  
QY 1942 gaatccctgttggcgaagacaacaaagatgcaatggaacccctaagaaggtctatgtct 2001  
|||||  
DB 502 gaatccctgttggcgaagacaacaaagatgcaatggaacccctaagaaggtctatgtct 561  
QY 2002 actgaagcacaataacagagaatgatacagctt 2034  
|||||  
DB 562 actgaagcacaataacagagaatgatacagctt 594

RESULT 11  
AAC79294  
ID AAC79294 standard; cDNA; 540 BP.  
XX  
AC AAC79294;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE Human lung tumour-specific cDNA #247.  
XX

WPI; 2000-638466/61.

The present sequence is given in a specification relating to compounds for therapy and diagnosis of lung cancer. Polypeptides comprising at least an immunogenic part of a lung tumor protein are disclosing the polypeptides are useful for inhibiting the development of cancer especially lung cancer. Samples of T cells expressing the polypeptides may be used to inhibit the development of cancer. The polypeptides are also used for detecting and monitoring the progression of cancer especially lung cancer.

Sequence 540 BP; 147 A; 123 C; 133 G; 136 T; 1 other

Query Match	8.5%;	Score 483.4;	DB 21;	Length 540;
Best Local Similarity	97.48;	Pred. No. 1.7e-122;		
Matches 555; Conservative	0			

QY	1990	agctctatgctactgaagcaataaagcaagcaataa	Indels	Gaps
			2;	4
			Mismatches	
			0;	

Db  
1 agtctatgtctactgaagcaataacgaaqaafratcaaatatcgcggcgaggatccagcatttggtgcaaggaga 2049

2050 ataagcaagtgcacaatgaactgaagtcaactggagccgccttgaacot-  
|||||

61 ataacgaagtgcatagtgctgaagtcacctgagagccccctgagaccctgaacttcccccatt 2109  
2110  
2111

121 gaacagcgttgatgatagagaagaattccattccctctacagtgtgattgac 3160  
|||||

Y 2170 ggqcttgaatcgcacccccc  
gaaacagcgcttgatgatagaagaacgaagtccattccctctacagtgsgattgag 180

181 ggscctgatgaatcgcgccacacgaatatggtgagtcaggt 2229

aaatacagctgcacctacagtgaatattaccaccccaaaa  
|||||  
gtt 231

232 aaataccagctgtccctacagtgataatgcccacaatnaccctat  
35cccccaagtagacactgtccattatagaagat 2289

2290 gacaggttcgcagtgcttccctcacatctctcgaacagtcgctcttccnagt+acc-  
||||| -gagacacacggcattatagaagat 291

**gacaggtgcgaagtcttcctccacaatcctcttaccaggacctctctccaqglcccatt**

gacagcgggtctgtgaaggcacatgcctgtacttggccaaggctgcaatcatgatcca 2409  
|||||  
352 gatgtccaaatt

[illegible]

412 gccgactgctcttgatccacatattgaatccattcttgcattcaagc-agaaggaatt 2468

...ccccccgcccgccttccaacgaagaaaggatt 471

RESULT	12
AAD23370	
ID	AAD23370 standard; cDNA; 540 BP.
XX	
CC	

DT  
XX  
26-FEB-2002 (first entry)

Human lung tumour-specific SCC1-568 cDNA.

antitumor protein; immunostimulant; cytostatic; gene therapy; vaccine; immune response; antitumor therapy; human; lung tumour

... response; Lung cancer; SCC1-568; ss

WO200172295-A2

04-OCT-2001.

28-MAR-2001; 2001WO-US09991

05-JUN-2000; 2000US-0538037  
18-AUG-2000; 2000US-0588937  
2000US-0540070

22-SEP-2000; 2000US-234517P.  
01-NOV-2000; 2000US-0704512  
14-DEC 2000; 2000US-0640878.

13 DEC-2000; 2000US-0738973.  
(CORI-) COPYA CORR

Reed SG, Lodes MJ.

Henderson RA, Fling SP, Algate PA, Elliot M, Mannion T, Benson DR, Indrias CV, WPT. 2001. 2001.

NEW HUMAN JUNE 2001 639201/73.

lung-specific polynucleotides and polypeptides for the diagnosis and treatment of disease e.g. lung cancer

Example 2; Page 276; 378pp; English

The invention relates to isolated lung tumour-specific proteins and their corresponding cDNA molecules. Lung tumour-specific proteins and expanding T cells specific for a tumour protein, and for inhibiting the development of cancer. The invention also relates to a composition useful for stimulating an immune response, and for treating cancer. Lung tumour specific oligonucleotide is useful in gene therapy and for diagnosis, detection and treatment of lung cancer. The present sequence is human lung tumour-specific cDNA.

sequence 540 BP; 147 A; 123 C; 133 G; 136 T; 1 other;

try Match  
t Local Similarity  
8.58; Score 482.4; nb 33

Category	Count	Percentage	Length
Conservative	535	97.4%	540
Mismatches	0	0.0%	122
Indels	2	0.4%	13

1990 aggtctatgtctactgaagccaataaacgaggaatgatccagctattgtttgcaagcaga 3040  
|||||  
1991 |||||  
1992 |||||  
1993 |||||  
1994 |||||  
1995 |||||  
1996 |||||  
1997 |||||  
1998 |||||  
1999 |||||  
2000 |||||  
2001 |||||  
2002 |||||  
2003 |||||  
2004 |||||  
2005 |||||  
2006 |||||  
2007 |||||  
2008 |||||  
2009 |||||  
2010 |||||  
2011 |||||  
2012 |||||  
2013 |||||  
2014 |||||  
2015 |||||  
2016 |||||  
2017 |||||  
2018 |||||  
2019 |||||  
2020 |||||  
2021 |||||  
2022 |||||  
2023 |||||  
2024 |||||  
2025 |||||  
2026 |||||  
2027 |||||  
2028 |||||  
2029 |||||  
2030 |||||  
2031 |||||  
2032 |||||  
2033 |||||  
2034 |||||  
2035 |||||  
2036 |||||  
2037 |||||  
2038 |||||  
2039 |||||  
2040 |||||  
2041 |||||  
2042 |||||  
2043 |||||  
2044 |||||  
2045 |||||  
2046 |||||  
2047 |||||  
2048 |||||  
2049 |||||  
2050 |||||  
2051 |||||  
2052 |||||  
2053 |||||  
2054 |||||  
2055 |||||  
2056 |||||  
2057 |||||  
2058 |||||  
2059 |||||  
2060 |||||  
2061 |||||  
2062 |||||  
2063 |||||  
2064 |||||  
2065 |||||  
2066 |||||  
2067 |||||  
2068 |||||  
2069 |||||  
2070 |||||  
2071 |||||  
2072 |||||  
2073 |||||  
2074 |||||  
2075 |||||  
2076 |||||  
2077 |||||  
2078 |||||  
2079 |||||  
2080 |||||  
2081 |||||  
2082 |||||  
2083 |||||  
2084 |||||  
2085 |||||  
2086 |||||  
2087 |||||  
2088 |||||  
2089 |||||  
2090 |||||  
2091 |||||  
2092 |||||  
2093 |||||  
2094 |||||  
2095 |||||  
2096 |||||  
2097 |||||  
2098 |||||  
2099 |||||  
2100 |||||  
2101 |||||  
2102 |||||  
2103 |||||  
2104 |||||  
2105 |||||  
2106 |||||  
2107 |||||  
2108 |||||  
2109 |||||  
2110 |||||  
2111 |||||  
2112 |||||  
2113 |||||  
2114 |||||  
2115 |||||  
2116 |||||  
2117 |||||  
2118 |||||  
2119 |||||  
2120 |||||  
2121 |||||  
2122 |||||  
2123 |||||  
2124 |||||  
2125 |||||  
2126 |||||  
2127 |||||  
2128 |||||  
2129 |||||  
2130 |||||  
2131 |||||  
2132 |||||  
2133 |||||  
2134 |||||  
2135 |||||  
2136 |||||  
2137 |||||  
2138 |||||  
2139 |||||  
2140 |||||  
2141 |||||  
2142 |||||  
2143 |||||  
2144 |||||  
2145 |||||  
2146 |||||  
2147 |||||  
2148 |||||  
2149 |||||  
2150 |||||  
2151 |||||  
2152 |||||  
2153 |||||  
2154 |||||  
2155 |||||  
2156 |||||  
2157 |||||  
2158 |||||  
2159 |||||  
2160 |||||  
2161 |||||  
2162 |||||  
2163 |||||  
2164 |||||  
2165 |||||  
2166 |||||  
2167 |||||  
2168 |||||  
2169 |||||  
2170 |||||  
2171 |||||  
2172 |||||  
2173 |||||  
2174 |||||  
2175 |||||  
2176 |||||  
2177 |||||  
2178 |||||  
2179 |||||  
2180 |||||  
2181 |||||  
2182 |||||  
2183 |||||  
2184 |||||  
2185 |||||  
2186 |||||  
2187 |||||  
2188 |||||  
2189 |||||  
2190 |||||  
2191 |||||  
2192 |||||  
2193 |||||  
2194 |||||  
2195 |||||  
2196 |||||  
2197 |||||  
2198 |||||  
2199 |||||  
2200 |||||  
2201 |||||  
2202 |||||  
2203 |||||  
2204 |||||  
2205 |||||  
2206 |||||  
2207 |||||  
2208 |||||  
2209 |||||  
2210 |||||  
2211 |||||  
2212 |||||  
2213 |||||  
2214 |||||  
2215 |||||  
2216 |||||  
2217 |||||  
2218 |||||  
2219 |||||  
2220 |||||  
2221 |||||  
2222 |||||  
2223 |||||  
2224 |||||  
2225 |||||  
2226 |||||  
2227 |||||  
2228 |||||  
2229 |||||  
2230 |||||  
2231 |||||  
2232 |||||  
2233 |||||  
2234 |||||  
2235 |||||  
2236 |||||  
2237 |||||  
2238 |||||  
2239 |||||  
2240 |||||  
2241 |||||  
2242 |||||  
2243 |||||  
2244 |||||  
2245 |||||  
2246 |||||  
2247 |||||  
2248 |||||  
2249 |||||  
2250 |||||  
2251 |||||  
2252 |||||  
2253 |||||  
2254 |||||  
2255 |||||  
2256 |||||  
2257 |||||  
2258 |||||  
2259 |||||  
2260 |||||  
2261 |||||  
2262 |||||  
2263 |||||  
2264 |||||  
2265 |||||  
2266 |||||  
2267 |||||  
2268 |||||  
2269 |||||  
2270 |||||  
2271 |||||  
2272 |||||  
2273 |||||  
2274 |||||  
2275 |||||  
2276 |||||  
2277 |||||  
2278 |||||  
2279 |||||  
2280 |||||  
2281 |||||  
2282 |||||  
2283 |||||  
2284 |||||  
2285 |||||  
2286 |||||  
2287 |||||  
2288 |||||  
2289 |||||  
2290 |||||  
2291 |||||  
2292 |||||  
2293 |||||  
2294 |||||  
2295 |||||  
2296 |||||  
2297 |||||  
2298 |||||  
2299 |||||  
2300 |||||  
2301 |||||  
2302 |||||  
2303 |||||  
2304 |||||  
2305 |||||  
2306 |||||  
2307 |||||  
2308 |||||  
2309 |||||  
2310 |||||  
2311 |||||  
2312 |||||  
2313 |||||  
2314 |||||  
2315 |||||  
2316 |||||  
2317 |||||  
2318 |||||  
2319 |||||  
2320 |||||  
2321 |||||  
2322 |||||  
2323 |||||  
2324 |||||  
2325 |||||  
2326 |||||  
2327 |||||  
2328 |||||  
2329 |||||  
2330 |||||  
2331 |||||  
2332 |||||  
2333 |||||  
2334 |||||  
2335 |||||  
2336 |||||  
2337 |||||  
2338 |||||  
2339 |||||  
2340 |||||  
2341 |||||  
2342 |||||  
2343 |||||  
2344 |||||  
2345 |||||  
2346 |||||  
2347 |||||  
2348 |||||  
2349 |||||  
2350 |||||  
2351 |||||  
2352 |||||  
2353 |||||  
2354 |||||  
2355 |||||  
2356 |||||  
2357 |||||  
2358 |||||  
2359 |||||  
2360 |||||  
2361 |||||  
2362 |||||  
2363 |||||  
2364 |||||  
2365 |||||  
2366 |||||  
2367 |||||  
2368 |||||  
2369 |||||  
2370 |||||  
2371 |||||  
2372 |||||  
2373 |||||  
2374 |||||  
2375 |||||  
2376 |||||  
2377 |||||  
2378 |||||  
2379 |||||  
2380 |||||  
2381 |||||  
2382 |||||  
2383 |||||  
2384 |||||  
2385 |||||  
2386 |||||  
2387 |||||  
2388 |||||  
2389 |||||  
2390 |||||  
2391 |||||  
2392 |||||  
2393 |||||  
2394 |||||  
2395 |||||

[illegible]

Db	61	ataagcaagtcgaataagctgtaagctaaccttggagagcccccctggaactgagctggcccaat	120
Qy	2110	gaacaacagcttggatgataagagaacgaagaatttccattccctctacagtgaggattgag	2169
Db	121	gaacaacagcttggatgataagagaacgaagaatttccattccctctacagtgaggattgag	180
Qy	2170	gggcttgaatgcgcgcgaacgaagaatgctgcgcctcagtagaatatggttgatcaggt	2229
Db	181	gggcttgatgaatgcgcgcgaacgaagaatgctgcgcctcagtagaataatg-----ggt	231
Qy	2220	aaataaccagctgtccctctaagatgaaatgagccccaagatgaacatgcatatagaagaat	2289
Db	232	aaataaccagctgtccctctaagatgaaatgagccccaagatgaacatgcatatagaagaat	291
Qy	2290	gacaggtgcgaatgctctctccacaatctctcgaaccagctcctctccagctcccatgat	2349
Db	292	gacaggtgcgaatgctctctccacaatctctcgaaccagctcctctccagctcccatgat	351
Qy	2350	gagtgtgggtttcttgacgacgacagatgctgtgtaacttggcccaaggtctgcaatcagtgata	2409
Db	352	gagtgtgggtttcttgacgacgacagatgctgtgtaacttggcccaaggtctgcaatcagtgata	411
Qy	2410	gcgcagctgcctttgtagtccagatgtgtgatacttccagtttctgtcttttcaacg-agaagagtt	2468
Db	412	gcgcagctgcctttgtagtccagatgtgtgatacttccagtttctgtcttttcaacgaaaaagagtt	471
Qy	2469	tggagcgtccag-agtatgtcagaaaaacgcga-caaagcaatttcagatgagccagtcgaattg	2526
Db	472	tggagcgtccagaaagtatgctcgaaaaaacgcgaccaaagcaatttcanaatgacgataatg	531
Qy	2527	gatttcgctt	2535
Db	532	gatttcgctt	540
RESULT 13			
ID	AAS68902	standard; cDNA; 510 bp.	
XX	AAS68902;		
AC	13-FEB-2002 (first entry)		
XX			
DT			
DE	DNA encoding novel human diagnostic protein #4706.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
PF	30-MAR-2001; 2001WO-US08631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Drmanac RT, Liu C, Tang YT;		
XX			
DR	WPI; 2001-639362/73.		
XX			
XX	P-PSDB; ABG04715.		
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity -		
XX			
XX	Claim 1; SEQ ID No 4706; 103bp; English.		
XX			

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. *AA564197-AA594564* represent novel human diagnostic coding sequences of the invention.  
 Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 510 BP; 133 A; 133 C; 144G; 100 T; 0 other;

XX DE Novel human polynucleotide, SEQ ID NO: 3345.  
 XX KW Human; cytosolic; gene therapy; colon cancer; prostate cancer;  
 XX KW breast cancer; lung cancer; cancer detection; ss.  
 OS Homo sapiens.  
 XX MO200102568-A2.  
 XX PN 11-JAN-2001.  
 XX PD 30-JUN-2000; 2000MO-US18374.  
 XX PF 02-JUL-1999; 99US-0142310.  
 XX PR 02-JUL-1999; 99US-0142311.  
 XX PA (CHIR) CHIRON CORP.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;  
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;  
 PI Cirenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;  
 XX Kita D, Garcia V, Jones LM, Strache-Crain B;  
 DR WPI: 2001-091805/10.  
 XX Library of polynucleotides for diagnosing a cancerous state of a  
 PT mammalian cell and detecting cancer, particularly of the colon or  
 XX prostate, comprises 3351 human polynucleotide sequences -  
 XX Claim 9; Page 1045; 1046pp; English.  
 CC The present sequence is one of 3351 sequences in a library of human  
 CC polynucleotides. The library is used to detect differentially expressed  
 CC genes correlated with the cancerous state of a mammalian cell and can  
 CC detect colon, prostate, breast and lung cancer. The library can be used  
 CC to produce probes for detection of mRNA and to produce additional copies  
 CC of the polynucleotides. The probes can be used for chromosome mapping of  
 CC the polynucleotide and for detection of transcription levels. Ribozymes  
 CC or antisense oligonucleotides can be generated. The polynucleotides and  
 CC their gene products are used as genetic or biochemical markers and  
 CC blood or tissues) that will detect the earliest changes along the  
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and  
 CC preventive interventions. The polynucleotides, polypeptides and  
 CC antibodies against them can be used in pharmaceutical compositions to  
 CC treat the cancers and proliferative disorders such as neoplasia, and  
 CC dysplasia and hyperplasia.  
 XX Sequence 425 BP; 134 A; 86 C; 110 G; 95 T; 0 other;

Query Match 73%; Score 415.4; DB 22; Length 425;  
 Best Local Similarity 99.8%; Pred. No. 4.7e-104;  
 Matches 416; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1732 ctacacatgagcagcaggaattctgacatttaagtcacactaattgtaga 1791  
 DB 9 ctacacatgagcagcaggaattctgacatttaagtcacactaattgtaga 1791  
 QY 1792 tctgagagcctgtgtcaggtcaaggtcaacoggtcaaaagagacacagcagattg 1851  
 DB 69 tctgagagcctgtgtcaggtcaaggtcaacoggtcaaaagagacacagcagattg 1851  
 QY 1852 ggaactctgcaagtcattatattgagagagcagcactcaaatggaagcttcgg 1911  
 DB 129 ggaactctgcaagtcattatattgagagagcagcactcaaatggaagcttcgg 1911  
 QY 1912 gtgatatcaactgataagcagtaaatggaagcactcctgttggcagacaacagaagt 1971  
 DB 189 gtgatatcaactgataagcagtaaatggaagcactcctgttggcagacaacagaagt 1971  
 QY 1972 gccatggaacccctaagaaggtctatgtctactgaaggcaataaagaggaattccag 2031

DB 249 gccatggaacccctaagaaggtctatgtctactgaaggcaataaagaggaattccag 308  
 QY 2032 ctattgttgcagaggaagataagcagatgcaatgagcgtgaagtcacctggagacccccc 308  
 DB 309 ctattgttgcagaggaagataagcagatgcaatgagcgtgaagtcacctggagacccccc 309  
 QY 2092 ggaactgagcctgcccatttgaacacggcttggatgataagagacgaagattcccat 2148  
 DB 369 ggaactgagcctgcccatttgaacacggcttggatgataagagacgaagattcccat 2148  
 RESTULT 15  
 ID AAH12111/c  
 XX AAH12111 standard; cDNA; 564 BP.  
 XX AC AAH12111;  
 XX DT 26-JUN-2001 (first entry)  
 XX DE Human CDNA clone (3'-primer) SEQ ID NO:8946.  
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX OS Homo sapiens.  
 XX PN EP1074617-A2.  
 XX PD 07-FEB-2001.  
 XX PF 28-JUL-2000; 2000EP-0116126.  
 XX PR 29-JUL-1999; 99JP-0248036.  
 XX PR 27-AUG-1999; 99JP-0300253.  
 XX PR 11-JAN-2000; 2000JP-0118776.  
 XX PR 02-MAY-2000; 2000JP-0183767.  
 XX PR 09-JUN-2000; 2000JP-0241899.  
 XX PA (HELI-) HELIX RES INST.  
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI: 2001-318749/34.  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 XX full-length cDNAs defined in the specification, and for the detection  
 XX and/or diagnosis of the abnormality of the proteins encoded by the  
 XX full-length cDNAs -  
 XX Claim 3; SEQ ID 8946; 2537bp + CD ROM; English.  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-gt primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH95893 represent human amino acid sequences; AAH92446 to  
 CC represent oligonucleotides, all of which are used in the exemplification



**us-09-757-781-20.rng**

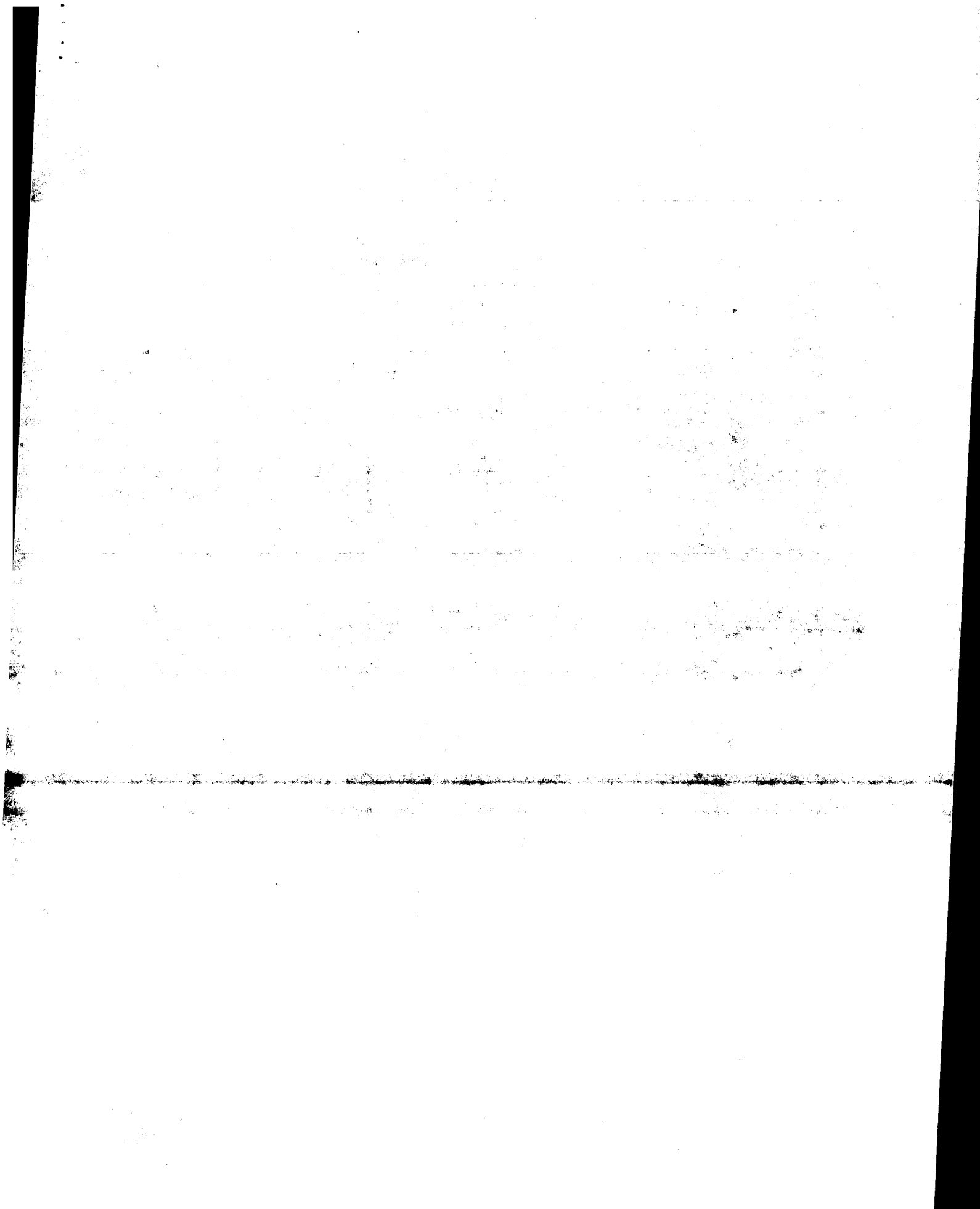
---

---

1.

587	
77	
47	
17	
807	
37	
667	
37	
227	
57	
887	
77	
247	
17	

1



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: July 24, 2002, 11:20:08 ; Search time 59.5 Seconds  
(without alignments)  
2531.362 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 7035  
Sequence: 1 MKVYVCFGRTRVVPCGDGH.....SQVRLRLRLQTPKGRPFYS 1356

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_032802:\*

1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4361	62.0	865	22	AA1995798 Human protein sequ
2	4361	62.0	22	AA1995798	Human protein sequ
3	3527	50.1	744	21	AA1995798 Human protein sequ
4	1397	19.9	273	22	AA1995798 Human protein sequ
5	1393	19.8	272	22	AA1995798 Human protein sequ
6	1134.5	16.4	1464	22	AA1995798 Human protein sequ
7	784	11.1	169	22	AA1995798 Human protein sequ
8	616	8.8	419	22	AA1995798 Human protein sequ
9	603.5	8.6	315	22	AA1995798 Human protein sequ
10	433.5	6.2	193	22	AA1995798 Human protein sequ
11	419.5	6.0	90	21	AA1995798 Human protein sequ

12	410	5.8	86	21	AA1995798
13	355.5	5.1	856	21	AA1995798
14	355.5	5.1	2037	21	AA1995798
15	355.5	5.1	2070	20	AA1995798
16	351	5.0	1373	20	AA1995798
17	351	5.0	2000	20	AA1995798
18	345.5	4.9	1005	20	AA1995798
19	343.5	4.9	763	20	AA1995798
20	340.5	4.8	1881	20	AA1995798
21	337.5	4.8	1526	22	AA1995798
22	337.5	4.8	1526	22	AA1995798
23	337.5	4.8	1526	22	AA1995798
24	337.5	4.8	1526	22	AA1995798
25	337.5	4.8	1554	22	AA1995798
26	337.5	4.8	1584	22	AA1995798
27	337.5	4.8	1627	22	AA1995798
28	333.5	4.7	1239	20	AA1995798
29	332	4.7	2466	16	AA1995798
30	332	4.7	2466	19	AA1995798
31	332	4.7	2466	21	AA1995798
32	324.5	4.6	2485	21	AA1995798
33	315.5	4.5	849	19	AA1995798
34	314.5	4.5	1743	22	AA1995798
35	307	4.4	60	22	AA1995798
36	307	4.4	60	22	AA1995798
37	305.5	4.3	318	20	AA1995798
38	301	4.3	1267	22	AA1995798
39	301	4.3	1267	22	AA1995798
40	294	4.2	590	20	AA1995798
41	293	4.2	1095	20	AA1995798
42	291	4.1	767	21	AA1995798
43	288	4.1	632	21	AA1995798
44	286	4.1	632	21	AA1995798
45	286	4.1	632	22	AA1995798

#### ALIGNMENTS

RESULT	ID	AA1995798	standard; Protein; 865 AA.
XX	AA1995798		
XX	26-JUN-2001	(first entry)	
DE	Human protein sequence SEQ	ID NO:18775.	
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy.		
XX	Homo sapiens.		
PN	EP1074617-A2.		
XX	07-FEB-2001.		
PF	28-JUL-2000; 2000EP-0116126.		
XX	29-JUL-1999; 99JP-0248036.		
PR	27-AUG-1999; 99JP-0300253.		
PR	11-JAN-2000; 2000JP-0118776.		
PR	02-MAY-2000; 2000JP-0183767.		
PR	09-JUN-2000; 2000JP-0241899.		
PA	(HELI-) HELIX RES INST.		
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
DR	WPI; 2001-318749/34.		
XX	Primer sets for synthesizing polynucleotides, particularly the 5602		

full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEQ ID 18775; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dt primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 865 AA:

Query Match 62.0%; Score 4361; DB 22; Length 865;

Best Local Similarity 94.8%; Pred. No. 3.1e-301; Matches 852; Conservative 5; Mismatches 8; Indels 34; Gaps 2;

QY 136 MFLHRRSSDPALIGLSTVSDSNFSESRKRNPTMTAFLKONTAGSPKTCDRKK 195  
Db 1 mplhrrssdpaliglstvsdsnfsestrknpctwtstgflqncatgspkcdrrk 60  
QY 196 DENRSLPRDTSNMSNOFORNARSSLASHPMYKWLKQEDDEGTEEDNRRVPGH 255  
Db 61 denyrrslprdtswnsnqqrharrsslsashpmgkwekqeqdedgteednrrvpgn 120  
QY 256 ADTGLEHTPNFSLDWMKLVKVEVPNDGPGPLGHVVPFSGRGTGLVYKRLKGGKAHE 315  
Db 121 adtglehtpnfslldwmkllvevpndgpgplghvvpfsargrqlglvkllegygkehe 180  
QY 316 NIFRNDQVIRINDGDLKRRREFQAHMFROAMRTPIIFHVHPVPAKNEQTEOLSOSEKN 375  
Db 181 nifrndqviriindgdlkrrrefeqahmfroamrtpiifhvhpvpaankeqteolsgsekn 240  
QY 376 NYSSNFSPDSQYIDNRSVNSAGLHTVQAPRLNHPPOIDSHSLPSPAPSKPPSPAP 435  
Db 241 nyssnfspdsqyidnrsvnsagltvqraprlnhppoidshslpssapkskppsap 300  
QY 436 ASAPQWVFSTVSSGNTFKKIKRLNIOLKGTGEGSITSROVYTCGSAPYYKNTLP 495  
Db 301 asapqvwfstvssgntfkikrlniolkgtgegsitsrovytcgsapyykntlp 360  
QY 496 RGAALDGRKLAGDRLEIENGVDLVGKSOEVEVYSLRSTKMGCTVSLVFEREDAFHPR 555  
Db 361 rgaalidgrklagdrleienvgdvlvgksoevevysllrstkmgtvslvferedafhp 420  
QY 556 LNAEPSONQIPKFKAKEDDVIYLPDGTREELTFEYPLNDGSGAGLVSKNRSKENHA 615  
Db 421 lnaepsonqipkfkakededviyldpdtreelfteyplndgsaglgvsknrskenha 480  
QY 616 DLGIFVKSIIINGAASKDGRLRVNDOLIAVNGSLLGKTRNODAMFELRSMREGKRG 675  
Db 481 dlgifvksiiingaaaskdgrlrvndoliavngsllgktrnodamfelfrsmregkr 540  
QY 676 IQLIVARRISKCNELKSPGSPPELPIETALDRERRISHSLYSGIEGLDESPRNAL 735

Db 541 Iqlivarriskcnelpspgsppepelpietaldreerrishsllysgiegldepsrnaal 600  
QY 736 SRIMESGKYQLSPYVMPODDTVIIEDRLPYLPRLHSDQSSSHDDVGFVTFADAGTW 795  
Db 601 srilm---gkyqlspvtvmpddtviieddrlpylprrlhdqssshddvgfvtadagtw 657  
QY 796 AKAAISDSADCSLSPDDVPYLAFOREGFGQSGMSKRTKOFSDASOLDPVTKRKSKMDL 855  
Db 658 akaaisdsadcsldpddvpylaforegfgqsgmskrtkofsdsoldpvtkrkskmdl 689  
QY 856 GIADETKINTVDDQKAGSPSRDVGPSGLKSSLSLELQTAVAEYTLNGDIPFRPRPRI 915  
Db 690 ---detkintvddqkagpsrdrvpslglksslselqtavaevtlngdiphrprpri 746  
QY 916 IYRGCSNESFRAIDKSYDKPAVDDDDGEMETLEEDTEESRSGRESVSTASDPSHLE 975  
Db 747 iyrgcsnesftraidksydkpavddddgmetleedteesrgresvstasdpshle 806  
QY 976 ROMNGOEKGDTRDKKDKTKGKRRKDDKDKKAKKGMKGLGDMFRGKRRKDDK 1034  
Db 807 romngoekgdtrdkdktkgkrrkddkdkkakkgmkgldmfrgkrrkddk 865

RESULT 2  
AA67610 standard; Protein; 865 AA.

AA67610:

26-NOV-2001 (first entry)

Amino acid sequence of a human protein kinase/protein phosphatase.  
Human; protein kinase; protein phosphatase; signal transduction.

Homo sapiens.

WO200109316-A1.

08-FEB-2001.

28-JUL-2000; 2000WO-JP05061.

29-JUL-1999; 99JP-0248036.

18-OCT-1999; 99US-0159590.

11-JAN-2000; 2000JP-0118776.

17-FEB-2000; 2000US-0183322.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;

Senoo C, Nezu J;

WPI; 2001-570286/64.

N-PDB; AAH78369.

New genes encoding proteins with protein kinase/protein phosphatase activity, useful in the diagnosis and treatment of diseases -

Claim 2; Page 76-82; 233pp; Japanese.

The present sequence represents human protein kinase/protein phosphatase. It is expected that the protein kinase/protein phosphatase gene participates in signal transduction in cells. The protein kinase/protein phosphatase polypeptides and polynucleotides are useful for developing diagnostics and treatment agents for human and animal diseases. The protein kinase/protein phosphatase polypeptides are useful as target molecules in designing novel drugs. The protein kinase/protein phosphatase polynucleotides are useful as a source of probes and

CC primers, which may be used to isolate homologous sequences.  
 XX Sequence 865 AA:

Query Match 62.0%; Score 4361; DB 22; Length 865;  
 Best Local Similarity 94.8%; Pred. No. 3.1e-301;  
 Matches 852; Conservative 5; Mismatches 8; Indels 34; Gaps 2;

136 MCLHVRSSDPAALIGLSTVSQSNFSESPSRKRNPTRMSTAGLTKONRAGSPKTDKK 195  
 |||  
 1 mphihrssdpaliglstvsqsnfseesprknprrwtaltagtlkqntagspkadirk 60

196 DENYRSLPDDTSNMWNOFQDNARSSLSASHPWGMKLEKQEDDEGTEEDNRPVGH 255  
 |||  
 61 denyrsldpdtntwmsnqfgrdnarslsashpmvqkqeqdeedgtednsrvepygh 120

256 ADTGLEHIDNFSLDNWKLEVPPNDGPGIGIHVPFSARGRTLLGLVRLKRGKRAHE 315  
 |||  
 121 adtglehlpnfslidmwnkivevpndgpglghvfpfsargrtllgllvkrlekgykaeh 180

316 NLPREDCIVRIINDGDLRRRREFQAOHMRPRTIMEHVVPANKRQYBQLSGSEKN 375  
 |||  
 181 nlprendcivrlindgdlrrrrefeqahmrtrqmrtpilwfhvpaankqeqqlsgsek 240

376 NYSSRFSFSDQYIDNRSYNASAGLHTVQAPRLNHPPEOIDSRSRLPHSAHPGKPPAP 435  
 |||  
 241 nyssrfsfspdsgyidnrsynasaglhvqraprlnhppegidshsrphsahpdkppsap 300

436 ASAPQWSTYVSSGYNTYKIGKRLNIQLKGTGEGFSTSDVTIGSAPLYVKNILP 495  
 |||  
 301 asapqwnstvtvssgyntykigkrlnlqlkgtgegfstsdvtilgsgaplyvknllp 360

496 RGAIDODGRLLKAGDRIEYNGVDLVGKSOEYVSLRSTRKMEGTVALYVRODAPHPRE 555  
 |||  
 361 rgaiddgrllkagdrilevngvdlvgkseeyvslrstckmevtsllyrtgdaahpre 420

556 LNAEPSQMOIPKRTKADEDDIVLTDPGTREFLFEVPLNDGSGAGLVGYVKNRSEKNA 615  
 |||  
 421 lnaepsqmipkrtkaadedivltpdgtreflfevplndsgsaglvgyvsknrsekh 480

616 DGIFFKSIINGAASAKDGRRLRYNDOLIANGSLIGKTQDMAMETLRKSMTEGKRM 675  
 |||  
 481 dgiiffksilingaasakdgrlryndqliavngesllgktqdametlrtsmstegnkrm 540

676 IOLIVARRISKCEKELKSPGSPGPELPIETALDRERIRSHLSYSGIEGIDESPRAAL 735  
 |||  
 541 iolivarriiskcelkspgspgpelpletaldrerirshlsysgiegidesprraal 600

736 SRIAGESGKYQLSPVTNMPDDTVIIEDDRPLVPLPHLSDQSSSSSHDDVGFYTADAGTW 795  
 |||  
 601 srim---gkyqlspvtnmpddtvlleddrplvplphlsdqssssshddvgyftadagtw 657

796 AKRAISDSADCSLSPVDVPLAFQRBGFGQNSSEKRTKOFSDAQLDVFYKTRKSKSMDL 855  
 |||  
 658 akaaisdsadcsldspvdvplafqrbgfgqnssekrkofsdlaqldvfyktrksksmdl 689

856 GIADETKLVMTVDQKAGSPRDVPSLGLKSSLSLEIOTAVAEWTLNGDIPHRPRRI 915  
 |||  
 690 ---deklntvdqkagsprdvpslglksslsleiqtaevetlningdiphrprri 746

916 IGRGGENESFRAAIDKSYDKPAVDDDEGMTLEEDTEESSRSGRESVSTASDQSHSL 975  
 |||  
 747 igrgenesfrraidsydkpavdddegmetleedteessrsgresvstasdqshsle 806

976 ROMNGNOEGDKTDKRRKKDKTGKREKKDKDROKEMKAKAKKMLGIGMFEFGKHAKDK 1034  
 |||  
 807 rgmngnqegdktdrrkkdktgkrekdkdrrdkemkakkakgmlkylgdmfslaklpekr 865

RESULT 3  
 AA71410  
 ID AA71410 standard; Protein; 744 AA.

XX AA71410;  
 AC  
 XX  
 DT 04-OCT-2000 (first entry)

DE PDZ domain-comprising PHIP (ephrlin Interacting Protein) complex.  
 DE  
 XX  
 XX PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;  
 XX cellular process; nerve cell interaction; regeneration of nerve cell;  
 KW axonogenesis; antiproliferative; proliferative disorder; treatment; GRIP;  
 KW differentiative disorder; PHIP; ephrin interacting protein; syntenin.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 273..351  
 FT Domain /label= PDZ\_domain  
 FT Domain 461..546  
 FT Domain /label= PDZ\_domain  
 FT Domain 590..678  
 FT Domain /label= PDZ\_domain

PN MO200031124-A2.  
 XX  
 PD 02-JUN-2000.  
 XX  
 PF 19-NOV-1999; 99WO-CA01101.  
 XX  
 PR 20-NOV-1998; 98US-0109158.  
 XX  
 PA (MOUN ) MOUNT SINAI HOSPITAL.  
 XX  
 PI Lin D, Pawson A;  
 XX  
 DR WPI: 2000-400038/34.  
 DR N-PSDB: AAD01182.  
 XX  
 XX  
 PT Isolated complex for treating proliferative or differentiative  
 PT disorders comprises B class ephrin and PDZ domain containing protein -  
 XX  
 PS Claim 26; Fig 2D; 59pp; English.

XX The patent discloses a complex comprising of a B class ephrin and PDZ  
 CC domain containing protein. B class ephrins function as ligands for Eph  
 CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a  
 CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ  
 CC binding site. This complex is used in methods to modulate the interaction  
 CC of a B class ephrin and PDZ domain containing protein and to identify  
 CC modulator compounds. It is also used for modulating cellular processes  
 CC like, axonogenesis, nerve cell interactions and regeneration of nerve  
 CC cells. The complex is also useful for treating proliferative or  
 CC differentiative disorders associated with this protein complex.  
 CC The present sequence is the PHIP (ephrlin Interacting Protein) complex,  
 CC comprising three PDZ domains. PHIP cDNA is isolated by screening the  
 CC lambdaExlox 10.5 day mouse embryo expression library. PHIP is closely  
 CC related to Caenorhabditis elegans PAR-3 protein, involved in regulating  
 CC polarity of the early embryo. PHIP sequence can be used as a probe, to  
 CC isolate other proteins like GRIP and syntenin.  
 CC  
 XX  
 SQ Sequence 744 AA:

Query Match 50.1%; Score 3527; DB 21; Length 744;  
 Best Local Similarity 91.4%; Pred. No. 5.1e-242;  
 Matches 679; Conservative 26; Mismatches 38; Indels 0; Gaps 0;

QY 1 MKYVCFGRTRVVVPCGDGHNKVFSLIQAVTRYRKAIAKDPNTWIOVHRLHEDGSLD 60  
 |||  
 1 mkyvfcgrtrvvvpcgdghnkvfsligqavtryrkavakdpmwyvhrlehdggild 60

DB 61 LDOTLGVADDDKRLVAVPEODPHHGDDGTSASSTGOSERIGSGELGNNVAPQPY 120  
 |||  
 61 lddllcvadddkrlvavpdeodphhggdgtssstgtselgsgelgntnvasatqpy 120

OY	121	ATSELEVP	PSVLR	RAMPL	PHVRSSD	PLIG	STSV	SNSS	NEE	PSK	NTR	HST	AGTL	180																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
Db	121	atsele	psv	lrampl	phvrssd	plig	stsv	snss	nee	psk	ntr	hst	agtl	180																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
OY	181	KONTAGS	PKTCD	RKKD	ENYRSI	PRD	TNMS	NOF	ORD	NARS	L	SASH	PMV	KWLE	EODE	240																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
Db	181	kntag	sptcd	rkkd	enyrsl	prd	tnms	nof	ord	nars	l	sash	pmv	kwle	eode	240																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
OY	241	DGTEED	NSR	VEPV	GHAD	GLEH	INE	S	LD	MK	LV	EP	ND	GG	EL	GHV	FFS	RG	RTIG	300																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Db	241	dgteed	nsr	vepv	ghad	gleh	ine	s	ld	mk	lv	ep	nd	gg	el	ghv	ffs	rg	rtig	300																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
OY	301	LLVKLE	EG	GA	KEH	EN	T	RE	ND	C	IV	R	IND	GL	R	R	RE	FO	A	OH	FR	OA	MP	PI	TF	W	VA	360																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
Db	301	llvkle	eg	ga	keh	en	t	re	nd	c	iv	r	ind	gl	r	r	re	fo	a	oh	fr	oa	mp	pi	tf	w	va	360																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
OY	361	ANKEQ	VE	LO	S	OSE	K	N	N	Y	S	S	R	F	S	D	S	Q	Y	I	D	N	R	S	V	S	A	G	L	T	V	O	R	A	P	L	N	H	P	P	O	I	D	S	H	S	R	420																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
Db	361	ankeq	ve	lo	s	ose	k	n	n	y	s	s	r	f	s	d	s	q	y	i	d	n	r	s	v	s	a	g	l	t	v	o	r	a	p	l	n	h	p	p	o	i	d	s	h	s	r	420																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
OY	421	LPHSN	H	PG	R	PS	K	P	S	A	P	A	S	P	O	V	E	T	T	Y	S	S	G	N	T	K	K	I	K	L	N	T	O	L	K	G	T	E	G	S	T	S	R	O	V	480																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
Db	421	lphsn	h	pg	r	ps	k	p	s	a	p	a	s	p	o	v	e	t	t	y	s	s	g	n	t	k	k	i	k	l	n	t	o	l	k	g	t	e	g	s	t	s	r	o	v	480																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
OY	481	TIGG	S	A	P	I	Y	V	N	N	I	P	R	G	A	I	O	D	G	R	L	K	A	G	D	R	L	E	V	N	G	V	D	L	G	K	O	E	E	V	S	L	N	R	T	K	K	E	G	T	V	540																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Db	481	tigg	s	a	p	i	y	v	n	n	i	p	r	g	a	i	o	d	g	r	l	k	a	g	d	r	l	e	v	n	g	v	d	l	g	k	e	e	v	s	l	n	r	t	k	k	e	g	t	v	540																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
OY	541	SLV	F	R	O	D	A	F	H	P	R	E	L	N	A	P	S	O	M	O	I	P	K	E	T	A	E	D	I	V	L	P	D	G	T	R	E	F	L	T	E	V	P	L	N	D	S	G	S	A	G	600																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Db	541	slv	f	r	o	d	a	f	h	p	r	e	l	n	a	p	s	o	m	o	i	p	k	e	t	a	e	d	i	v	l	p	d	g	t	r	e	f	l	t	e	v	p	l	n	d	s	g	s	a	g	600																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
OY	601	LGVS	T	K	G	R	S	K	R	S	E	N	A	D	G	I	P	Y	K	S	I	N	G	A	S	K	O	G	R	L	R	V	N	D	L	A	V	N	G	E	S	L	I	K	T	N	D	A	N	E	660																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Db	601	lgvs	t	k	g	r	s	k	r	s	e	n	a	d	g	i	p	y	k	s	i	n	g	a	s	k	o	g	r	l	r	v	n	d	l	a	v	n	g	e	s	l	i	k	t	n	d	a	n	e	660																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
OY	661	TLR	S	M	S	T	E	G	K	R	O	M	I	L	I	V	A	R	I	S	K	O	N	E	L	S	K	S	P	G	P	P	E	L	P	I	E	T	A	L	D	D	E	R	I	S	I	S	T	S	720																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Db	661	tlrsm	s	t	e	g	k	r	m	i	l	i	v	a	r	i	s	k	o	n	e	l	s	k	s	p	g	p	p	e	l	p	i	e	t	a	l	d	d	e	r	i	s	i	s	t	s	720																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
OY	721	GIEG	LD	ES	PS	R	N	A	L	S	R	I	M	E	S	G	743																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																

PF	7-JAN-2001	2001MO-US01312
XX	31-JAN-2000	2000US-0179065
PR	04-FEB-2000	2000US-0180628
PR	24-FEB-2000	2000US-0184664
PR	02-MAR-2000	2000US-0186350
PR	16-MAR-2000	2000US-0189874
PR	17-MAR-2000	2000US-0190076
PR	18-APR-2000	2000US-0196423
PR	19-MAY-2000	2000US-0205515
PR	07-JUN-2000	2000US-0209467
PR	28-JUN-2000	2000US-0214886
PR	30-JUN-2000	2000US-0215135
PR	07-JUL-2000	2000US-0216647
PR	11-JUL-2000	2000US-0216880
PR	11-JUL-2000	2000US-0217487
PR	11-JUL-2000	2000US-0217496
PR	14-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225213
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225257
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	22-AUG-2000	2000US-0226261
PR	22-AUG-2000	2000US-0226268
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227182
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0233080
PR	12-SEP-2000	2000US-0233081
PR	12-SEP-2000	2000US-0233198
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0234400
PR	14-SEP-2000	2000US-0234401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0234997
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369

PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCT INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX MPI; 2001-465460/50.  
XX  
XX N-PSDB; AAS27320.  
XX

PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
PT prognosing disorders related to the proteins, including cancers, immune  
PS disorders and neuronal disorders -  
XX  
XX Claim 1: SEQ ID No 968; 880pp; English.  
XX  
CC The invention relates to novel isolated polypeptides (I), and  
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
CC diagnosing, preventing and treating diseases including immune system  
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
CC transplant rejections and graft versus host disease, infectious diseases  
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
CC other blood-related disorders (sickle cell anaemia), myeloproliferative  
CC disorders, primary haematopoietic disorders, hyperproliferative  
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal  
CC abnormalities (Down syndrome), Ischaemic Injury (e.g. stroke), renal  
CC disorders (e.g. glomerulonephritis), cardiovascular disorders  
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in  
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.  
CC Addison's disease), reproductive system disorders, gastrointestinal  
CC disorder (inflammatory disorders), liver disorders (cirrhosis),  
CC as stimulators of B-cell responsiveness to pathogens, activators of  
CC T-cells, to induce higher affinity antibodies, and as a means to induce  
CC tumour proliferation in pathologies e.g. acquired immune deficiency  
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction  
CC pathway protein, amino acid sequences of the invention.

Query Match 19.9%; Score 1397; DB 22; Length 273;  
Best Local Similarity 99.3%; Pred. No. 3, 4e-91;  
Matches 270; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 260 LEHINPESLDDMKYKLYEVNDGSLGIHYVPFARAGRTIGLYKRLKCKRAEHENLFR 319  
DB 2 LEHINPESLDDMKYKLYEVNDGSLGIHYVPFARAGRTIGLYKRLKCKRAEHENLFR 61  
QY 320 ENDCIYRINDGDLRNRFRFOAOHMFROAMRTPIITWVHPAANKEDYEOLOSEKNNYYS 379  
DB 62 ENXCIVRINDGDLRNRFRFOAOHMFROAMRTPIITWVHPAANKEDYEQISGEKNNYYS 121  
QY 380 SREPSDQYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHSRPLSHASHPGKPSAPASAP 439  
DB 122 SRTSPDSQYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHSRPLSHASHPGKPSAPASAP 181  
QY 440 QNVFTTYSAGYTKKIKGRKLNQLKKGTEGLFSITTSRYVTGGSPRIYKNIILRGAA 499  
DB 182 QNVFTTYSAGYTKKIKGRKLNQLKKGTEGLFSITTSRYVTGGSPRIYKNIILRGAA 241  
QY 500 IODGRKAGDRLIEVNGVPLVGSQSEWVSL 531  
DB 242 IQDGRKAGDRLIEVNGVPLVGSQSEWVSL 273

RESULT 5  
AAG74163  
ID AAG74163 standard; Protein; 272 AA.  
XX  
XX AAG74163;  
XX  
XX 03-SEP-2001 (first entry)  
XX  
XX Human colon cancer antigen protein SEQ ID NO:4927.  
XX  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX colorectal carcinoma.  
XX  
XX Homo sapiens.  
XX  
XX OS  
XX PN WO200122920-A2.  
XX  
XX PD 05-APR-2001.

XX 28-SEP-2000; 2000MO-US26524.  
 PF  
 XX  
 PR 29-SEP-1999; 99US-0157137.  
 PR 03-NOV-1999; 99US-0163280.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Barash SC, Birse CE, Rosen CA.  
 DR WPI: 2001-235357/24.  
 DR N-PSDB: AAH33594.  
 XX  
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 PT useful for preventing, diagnosing and/or treating colorectal cancers -  
 PS Claim 11: Page 6691-6693; 9803pp; English.  
 XX  
 CC AAH32943 to AAH37195 and AAH373514 to AAH37788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patients own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated P,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC present invention.  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
 XX  
 SO Sequence 272 AA;

Query Match 19.8%; Score 1393; DB 22; Length 272;  
 Best Local Similarity 99.3%; Pred. No. 6.5e-91;  
 Matches 269; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 260 LEHITPESLIDDMVKLVEVPNDGPGPLIHVVPFSAKGRGLGLVKKLEKGAHEHNEFR 319  
 DB 2 LEHITPESLIDDMVKLVEVPNDGPGPLIHVVPFSAKGRGLGLVKKLEKGAHEHNEFR 319  
 QY 320 ENDICIVRINDGDLNRRFEQAOHMFROAMRTPIIMFHVVPANKEOYQOLSQSEKNYYS 379  
 DB 62 ENXCIVRLNDGDLNRRFEQAOHMFROAMRTPIIMFHVVPANKEOYQOLSQSEKNYYS 379  
 QY 380 SRFSPSOYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHSLRPHSAHPSGKPSAPASAP 439  
 DB 122 SRFSPSOYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHSLRPHSAHPSGKPSAPASAP 439  
 QY 440 QNWFSTVSSGYNTKTKIGRLNIOLKKTBEGLGFSITSRDVTGGSAPIYVKMIILPRGAA 499  
 DB 182 QNWFSTVSSGYNTKTKIGRLNIOLKKTBEGLGFSITSRDVTGGSAPIYVKMIILPRGAA 499  
 QY 500 IDGRLKADRLIEVNGDVLVKSQEEVSL 530  
 DB 242 IDGRLKADRLIEVNGDVLVKSQEEVSL 530

RESULT 6  
 ABB61199 standard; Protein; 1464 AA.  
 ID ABB61199; AC ABB61199;  
 XX 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 10389.  
 XX KW Drosophila: developmental biology; cell signalling; insecticide;  
 XX KW pharmaceutical.  
 XX OS Drosophila melanogaster.  
 XX PN WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001MO-US09231.  
 XX PR 23-MAR-2000; 2000US-191637P.  
 XX PR 11-JUL-2000; 2000US-0614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX DR WPI: 2001-656860/75.  
 XX DR N-PSDB: ABL05302.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Disclosure; SEQ ID NO 10389; 21pp + Sequence listing; English.  
 CC  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116175) and the encoded proteins  
 CC sequences (AB116176-AB130511), expressed DNA  
 CC (AB130512-AB130513).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIP0  
 CC at ftp://wipo.int/pub/published\_pct\_sequences.  
 XX  
 SO Sequence 1464 AA;

Query Match 16.4%; Score 1154.5; DB 22; Length 1464;  
 Best Local Similarity 26.2%; Pred. No. 7.1e-73;  
 Matches 433; Conservative 211; Mismatches 470; Indels 537; Gaps 61;

QY 1 MKTVCGGRTRVYVPCGDGSHKVFSLIOAVTRYRKALADPNYVQVHRLHGDGDIID 60  
 DB 1 MKTVCGGRTRVYVPCGDGSHKVFSLIOAVTRYRKALADPNYVQVHRLHGDGDIID 60  
 QY 61 LDDIICDVAADKRLVAVFEEDQ-----PHHGDDGTS-ASSTGOSPELSELTNNVS 114  
 DB 59 LDDIICDVAADKRLVAVFEEDQ-----PHHGDDGTS-ASSTGOSPELSELTNNVS 114  
 QY 115 APOPYQ-ATSELEWTPPS-----VLRANMPLHVRSSDPALIGSTVSNSFSSEPPSRKN 169  
 DB 117 PCPCPDSTPHLEVLTSTSGPMAGLGVLMVRRSSDPALIGSTVSNSFSSEPPSRKN 169  
 QY 170 PTKRSTTAGF-----LKONTAGSPKTCRKADEN-----YSLPNDUSNNMS 210  
 DB 166 -krwsaaaphyagdsperlfldkagggqlspqweeddpsbqlkeqllhqgqhaangss 224  
 QY 211 NQFQRDNARSLSASHPMVG-----KMLEKOE--ODEGTEE----- 245  
 DB 225 ssgnqhpfarsgrlsmqlllgdngykwmeaeklqnppeqqlvyygqshahgngngays 284  
 QY 246 -----DNSRVYVGHADTGLEHINPFSLDMMVKLVVPPNDGGPLGIVHVPFSAKGRGL 299  
 DB 285 sks]preskrkeplgagayesire-----kdgenllllneygspgl]talpkengng--- 335



QY	300	GLLVKRLREKGGKKAHEHMLPFRENDQCYLRINDGDLRRRFRQAOHMRQAMRPIIWEHVP	359
Db	336	gllvqhhepsrregrl-rdrdlleingikligiteeqveqlrle-	384
QY	360	AANKEQXEOLOSCKNNVYSRFRSPSOYIDNNSVSAQLHVQAPRLNHPEQIDSHS	419
Db	385	--ssellvrrllrgdrn-----rqgrdskvaemveavts-----	417
QY	420	RLPHSAHPGKPPSPAPASABQNVSTTVSSGVYTKTKIGKRLNIOLKGTGEGISITSRD	479
Db	418	-----ptkrphnaprgvtalgv-----antxrlgkietimlkkpgnvgfvtlrd	462
QY	480	VTIGCSAPITYVKNIPLPGGAILDGRKAKGDRLEIYNGVDLVKSOEDEVSLIRTKMEGT	539
Db	463	npagqhepcriyiknllprgaaleiedgrlpgdrillevgdtpmtgktctdvallrgmpagat	522
QY	540	VSLLVFPO-----EDAFHPRRLN-----EFS-----	561
Db	523	vrlvsvrqqqlaeqadqppakssagvavavappapaaapappipvqkssarslftb	582
QY	562	---QWQIRK-----ETKAEDIEDVLPDPG---TRFEPLTFPVPLNDSSGAGIAGSVKG-	607
Db	583	qgqgqlnesqhtfidagesaasndslppssnshsrteeltlhlpyhdcakgllgsvtyvk	642
QY	608	-----NRSKENHADGIFVCKSTINGGAASKDGRLRVNDOLIAVNGESLL	651
Db	643	tcsnlnaagsaasnglmkhdgdlfykvnhlbgaastrdgrlrmddqllsvngvslr	702
QY	652	GKTQMDAELLRRSM-STEBNKRGMQLIYVARISNC-----	687
Db	703	qgnaaemelttrawmvtltpkhhpgtlltlyvgrllirsasgdlldhnsbshsnasgq	762
QY	688	-----NELKSPGSPPELPIETALDDRRRIRSHSLYGIIGLDESPRNALRI	738
Db	763	snsgsgnmnnssnaadnsgatvlylspckeqrcngggsgsgnemtmsnprldrl	822
QY	739	MG-----ESGKQLSPVYNMPQ-----	755
Db	823	tgylcsnaspsqsbqhqphpsqqqqrllpaaprcsaalrnesymatindwsp	882
QY	756	-----DDVPIIEDDLPLVLP-----PHLSDDSSSSSHD-----DVGCV	788
Db	883	aqgmhlmtahqgtllalleddeemprclparphqgchncslaspsqnlavngnpprlntv	942
QY	789	TADAGTWAKAALISDSADCSIPVDVPLAFQREGFQSGMSEK-----RTK	834
Db	943	pgtrpstsntfatysqslsietn-sgyehstraldjrrtsisekhaaaladatetcygtnk	1001
QY	835	QFSDASOLD-FVKTRKS-----KSMDLGID------ETKLN-T	865
Db	1002	klreererrllqtktaavygsgslstatarlasanaqsfgykhaktassleqretqqla	1061
QY	866	VDDQKAGSPNDVPSLGLKSSSLBSLQYAAVEVTLNDDIPRHRPRPILINGRCNSF	925
Db	1062	aaeeaaedqldgprslgmksslslelqmwqelms--degrbhqalrarpgrjreds1	1120
QY	926	RAALDKSYD--KP-----AVDDEBGMETLEEDPEESRSRGSRESVSTASDQPSHSLEROMN	979
Db	1121	raavvsopaaskprkwtllldgd-----hegsfaagtrng-----pfgsals1n	1161
QY	980	GNOEGDKDTRKKDKDTGKEKKKDKDKEKKDKAKKMLGSLGDMFEGKHRRDKTIERTG	1039
Db	1201	yavnlspptvavslatprqlqqgq-----qqlqghbqgq-----	1235
QY	1097	MYGGSVSEGSMLANRPOSREGCHMDALYAQVKRKPRNSKSPVDSNKSSTSN-----	1150
Db	1236	-----qlptaa--laaelerngkppayqpppplpapnrgvsgnllhqd	1275
QY	1151	-----HDIRQLRLQDF--QAKQKODEVEDRRRTYSFEQPPPNAR	1188

Db 1276 ifnhryqhyanyedlhqgbqhqqlstrhghyhsgrsarsqdv-smhsstssgsgpyslasp 1334  
QY 1189 AROS-GRHSVS-----VEVOMQROROBERRSQOAAQOYSSLPRQSRKNSASVSQDSWE- 1241  
Db 1335 gqgsgngvrpmssyryetvrgqgrvgakshhsaa-----tssssplnphwka 1384  
QY 1242 ---QNTSPGEGFQSAKENPRTSYSGSRNGYLGCHGFNARVMELETOELLROE----- 1290  
Db 1385 aamngyspsplnsasar-----srqfvf-----tytliregsgsglpah 1422  
QY 1291 --QRKEQOMKQPPS-----EGSSNTDS 1312  
Db 1423 l1qgbqgqqlqgqgpcyqlvqkmspsqygs 1453  
RESULT 7  
ABG04715  
ID ABG04715 standard; Protein; 169 AA.  
XX  
AC ABG04715;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #4706.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PM Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB; AAS68902.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 35074; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 169 AA;

Query Match 11.1%; Score 784; DB 22; Length 169;  
Best Local Similarity 93.8%; Pred. No. 6.5e-48;  
Matches 150; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1068 REFERRQARENDYAEIOTFHRTGCDDELMYGVSYEGSMALNARPOSPREGHMDALY 1127

Db 9 frefrqrerdyaellyfhtitgcdehmgdsyjdramalnarpqspregimmdaly 68

OY 1128 AOVKKPRNSKSPVDNSRTPSNHDIRIQRLRQEFQAKODEVDYDRRTYSFEQPPNAR 1187

Db 69 aqvkkprnkspsvdtntstpsnhdrigrirgelfqakqdedvdrtrlysfegppnar 128

OY 1188 PATQSGRHSVSEVOMQROREFRESSQOARQOXSILPRQ 1227

Db 129 pathsgrhshsvsevgmhrgqrgeesssqaqdqrqysilprq 168

RESULT 8

ABG04716  
ID ABG04716 standard; Protein; 419 AA.

XX ABG04716;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #4707.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS68903.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 20; SEQ ID NO 35075; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

XX and gene mapping, and in recombinant production of (II). The

XX polynucleotides are also used in diagnostics as expressed sequence tags

XX for identifying expressed genes. (I) is useful in gene therapy techniques

XX to restore normal activity of (II) or to treat disease states involving

XX a food supplement. (II) and its binding partners are useful for treating

XX disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG030377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 419 AA;

Query Match 8.8%; Score 616; DB 22; Length 419;

Best Local Similarity 71.5%; Pred. No. 2.2e-35;  
Matches 123; Conservative 4; Mismatches 17; Indels 28; Gaps 1;

OY 1026 FCKHRRDKIEKTKIKIQESFTSEERIRNKQOEFLQAKTRREFRQARENDYAEIOT 1085

Db 113 fgkhrkdkiektykikigesftseerlrnkqgegrlqaktrfrefrqrerdyaelgd 172

OY 1086 FHRTEGCDDELMYGVSYEGSMALNARPOSPREGHMDALYAOVKKPRNSKSPVDNSR 1145

Db 173 fhrtgcdehmgdsyjdramalnarpqspregimmdalyaovkkprnkspsvdkg 232

OY 1146 STPSNHDRIQRLRQEFQAKODEVDYDRRTYSFEQPPNARPATQSGRHSV 1197

Db 233 kvlgeadpv-----yprnptegnasl 256

RESULT 9

ABB1149  
ID ABB1149 standard; peptide; 315 AA.

XX ABB1149;

DT 11-JAN-2002 (first entry)

DE Human PAR3 homologue, SEQ ID NO:1519.

XX Human: cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibin; chemotaxis; chemokinesis; cancer; tumour; haematopoietic disorder;  
KW proliferation; metastasis; lymphoid cell disorder; asthma; arthritis;  
KW myeloid cell disorder; lymphoid cell disorder; proliferative retinopathy;

KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;

KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;

KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
KW antifungal; vulnery; antifolcer.

XX Homo sapiens.

OS WO200157188-A2.

PD 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US03800.

PR 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PA (HYSE-) HYSQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-457740/49.

DR N-PSDB; ABA08393.

PT Human proteins and DNA encoding sequences useful for preventing,



```
RESULT 11
AA71424
ID AAY71424 standard; Protein; 90 AA.
XX
AC AAY71424;
XX
DT 04-OCT-2000 (first entry)
XX
DE PDZ domain-comprising PHIP PDZ3 protein complex.
XX
KW PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
KW cellular process; nerve cell interaction; regeneration of nerve cell;
KW axonogenesis; antiproliferative; proliferative disorder; treatment; GRIP;
KW differentiative disorder; PHIP PDZ3 complex; ephrin interacting protein.
XX
OS Mus sp.
XX
PN WO200031124-A2.
XX
PD 02-JUN-2000.
XX
PF 19-NOV-1999; 99WO-CA01101.
XX
PR 20-NOV-1998; 98US-0109158.
XX
PA (MOUN ) MOUNT SINAI HOSPITAL.
XX
PI Lin D, Pawson A;
XX
DR WPI; 2000-400038/34.
XX
PT Isolated complex for treating proliferative or differentiative
XX disorders comprises B class ephrin and PDZ domain containing protein -
XX
PS Claim 3; Fig 2C; 59pp; English.
XX
CC The patent discloses a complex comprising of a B class ephrin and PDZ
CC domain containing protein. B class ephrins function as ligands for Eph
CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a
CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
CC binding site. This complex is used in methods to modulate the interaction
CC of a B class ephrin and PDZ domain containing protein and to identify
CC modulator compounds. It is also used for modulating cellular processes
CC like, axonogenesis, nerve cell interactions and regeneration of nerve
CC cells. The complex is also useful for treating proliferative or
CC differentiative disorders associated with this protein complex.
CC The present sequence is the PDZ domain-comprising PHIP (ephrin
CC interacting protein) PDZ3 protein complex. PHIP consists of three PDZ
CC domains and is closely related to Caenorhabditis elegans PAR-3 protein,
CC involved in regulating polarity of the early embryo. This sequence
CC is isolated by screening the lambdaExlox 10.5 day mouse embryo
CC expression library.
XX
SQ Sequence 90 AA;
XX
XX
Query Match 6.0%; Score 419.5; DB 21; Length 90;
Best Local Similarity 95.6%; Pred. No. 2.1e-22;
Matches 86; Conservative 2; Mismatches 1; Indels 1; Gaps 1;
QY 590 EYP-LNDSGSAGLGVSVKGRSKENRNLGIFPKSITNGSGASKDGLRYNDOLIVNGE 648
DB 1 eypLndsgsaglgvsvkgrskenhadlgifkxstlngsgaaskdglryndqlavng 60
QY 649 SLGKTNQDAMETLRSMSTEGNKGMIQL 678
DB 61 slgkangameeltrsmstegnkrgmigl 90
RESULT 12
AA71423
ID AAY71423 standard; Protein; 86 AA.
```

```
XX
AC AAY71423;
XX
DT 04-OCT-2000 (first entry)
XX
DE PDZ domain-comprising PHIP PDZ2 protein complex.
XX
KW PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
KW cellular process; nerve cell interaction; regeneration of nerve cell;
KW axonogenesis; antiproliferative; proliferative disorder; treatment;
KW differentiative disorder; PHIP PDZ2 complex; ephrin interacting protein.
XX
OS Mus sp.
XX
PN WO200031124-A2.
XX
PD 02-JUN-2000.
XX
PF 19-NOV-1999; 99WO-CA01101.
XX
PR 20-NOV-1998; 98US-0109158.
XX
PA (MOUN ) MOUNT SINAI HOSPITAL.
XX
PI Lin D, Pawson A;
XX
DR WPI; 2000-400038/34.
XX
PT Isolated complex for treating proliferative or differentiative
XX disorders comprises B class ephrin and PDZ domain containing protein -
XX
PS Claim 3; Fig 2C; 59pp; English.
XX
CC The patent discloses a complex comprising of a B class ephrin and PDZ
CC domain containing protein. B class ephrins function as ligands for Eph
CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a
CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
CC binding site. This complex is used in methods to modulate the interaction
CC of a B class ephrin and PDZ domain containing protein and to identify
CC modulator compounds. It is also used for modulating cellular processes
CC like, axonogenesis, nerve cell interactions and regeneration of nerve
CC cells. The complex is also useful for treating proliferative or
CC differentiative disorders associated with this protein complex.
CC The present sequence is the PDZ domain-comprising PHIP (ephrin
CC interacting protein) PDZ2 protein complex. PHIP consists of three PDZ
CC domains and is closely related to Caenorhabditis elegans PAR-3 protein,
CC involved in regulating polarity of the early embryo. This sequence
CC is isolated by screening the lambdaExlox 10.5 day mouse embryo
CC expression library.
XX
SQ Sequence 86 AA;
XX
XX
Query Match 5.8%; Score 410; DB 21; Length 86;
Best Local Similarity 97.7%; Pred. No. 9.5e-22;
Matches 84; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 461 NIQLKKTEGLGFSITRDVTIGSGAPYIVKNIPLPGAAIOGRIKAGDRILEVNGVDLV 520
DB 1 nqlkkteglgfsitrdvtlgsasapyivknllprgaalgdrilkaqdrilevngvda 60
QY 521 GKSQEEVSLNKRSTKMEGVSLVFR 546
DB 61 gksqeevslnkrstkmegvslvfr 86
RESULT 13
AAB01383
ID AAB01383 standard; Protein; 856 AA.
XX
AC AAB01383;
XX
DT 20-OCT-2000 (first entry)
```

XX DE Neuron-associated protein.

KW KW Neuron associated protein; NEUP; neurological disorder; epilepsy;  
 KW KW Ischemic cerebrovascular disease; stroke; cerebral neoplasm;  
 KW KW Alzheimer's disease; Pick's disease; Huntington's disease;  
 KW KW dementia; Parkinson's disease; demyelinating disease; meningitis;  
 KW KW prion disease; Kuru; Creutzfeldt-Jakob disease; neurofibromatosis;  
 KW KW cerebral palsy; muscular dystrophy; central nervous system; CNS;  
 KW KW peripheral nervous system; PNS; myopathy; schizophrenia;  
 KW KW actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;  
 KW KW cirrhosis; hepatitis; mixed connective tissue disease; MCTD;  
 KW KW myelofibrosis; paroxysmal nocturnal hemoglobinuria; cancer;  
 KW KW autoimmune disease; inflammation; acquired immunodeficiency syndrome;  
 KW KW AIDS; Addison's disease; adult respiratory distress syndrome;  
 KW KW allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;  
 KW KW Werner syndrome; trauma; human.

XX OS Homo sapiens.

XX FH Key

FT FT Domain

FT FT Location/Qualifiers

FT FT 1..65

FT FT /label= PDZ domain signature

FT FT 18

FT FT /note= "Potential glycosylation site"

FT FT 20

FT FT /note= "Potential phosphorylation site"

FT FT 69

FT FT /note= "Potential phosphorylation site"

FT FT 71

FT FT /note= "Potential phosphorylation site"

FT FT 73

FT FT /note= "Potential phosphorylation site"

FT FT 118

FT FT /note= "Potential phosphorylation site"

FT FT 119

FT FT /note= "Potential phosphorylation site"

FT FT /note= "Potential phosphorylation site"

FT FT 136..218

FT FT /label= PDZ domain signature

FT FT 144..147

FT FT /note= "Glycosaminoglycan attachment site"

FT FT 199

FT FT /note= "Potential glycosylation site"

FT FT 238

FT FT /note= "Potential phosphorylation site"

FT FT 253

FT FT /note= "Potential phosphorylation site"

FT FT 263

FT FT /note= "Potential phosphorylation site"

FT FT 269..349

FT FT /label= PDZ domain signature

FT FT 284

FT FT /note= "Potential phosphorylation site"

FT FT 296

FT FT /note= "Potential phosphorylation site"

FT FT 337

FT FT /note= "Potential phosphorylation site"

FT FT 341

FT FT /note= "Potential phosphorylation site"

FT FT 369

FT FT /note= "Potential phosphorylation site"

FT FT 370

FT FT /note= "Potential glycosylation site"

FT FT 386

FT FT /note= "Potential phosphorylation site"

FT FT 389

FT FT /note= "Potential phosphorylation site"

FT FT 390

FT FT /note= "Potential glycosylation site"

FT FT 414

FT FT /note= "Potential phosphorylation site"

FT FT 415..497

FT FT /label= PDZ domain signature

FT FT Modified-site

FT FT 475

FT FT /note= "Potential phosphorylation site"

FT FT 511..592

FT FT /label= PDZ domain signature

FT FT 531

FT FT /note= "Potential glycosylation site"

FT FT 568

FT FT /note= "Potential glycosylation site"

FT FT 599

FT FT /note= "Potential phosphorylation site"

FT FT 604

FT FT /note= "Potential phosphorylation site"

FT FT 625

FT FT /note= "Potential phosphorylation site"

FT FT 633

FT FT /note= "Potential phosphorylation site"

FT FT 634

FT FT /note= "Potential phosphorylation site"

FT FT 648..733

FT FT /label= PDZ domain signature

FT FT 690

FT FT /note= "Potential phosphorylation site"

FT FT 705

FT FT /note= "Potential phosphorylation site"

FT FT 721

FT FT /note= "Potential glycosylation site"

FT FT 773..856

FT FT /label= PDZ domain signature

FT FT 819..821

FT FT /label= Cell attachment sequence

FT FT 835

FT FT /note= "Potential phosphorylation site"

XX PN MO200034477-A2.

XX PD 15-JUN-2000.

XX PF 10-DEC-1999; 99MO-US30408.

XX PR 11-DEC-1998; 98US-0210083.

XX PR 11-DEC-1998; 98US-9123456.

XX PR 09-FEB-1999; 99US-0119365.

XX PR 16-MAR-1999; 99US-0124687.

XX PA (INCY-) INCYTE PHARM INC.

XX PI Tang YT, Yue H, Baughn MR, Hillman JL, Lal P, Au-Young J, Yang J;

XX PI Lu DAM, Azimzai Y;

XX DR WPI; 2000-423423/36.

XX DR N-PSDB; AAA47424.

XX PS New human neuron-associated proteins and polynucleotides encoding them,  
 PT useful for diagnosis, treatment and prevention of cell proliferative  
 PT disorders including cancer, neuronal and neurological disorders

XX Claim 1; Page 105-108; 145pp; English.

CC CC Human neuron-associated proteins (NEUP) can be used for for  
 CC treating or preventing a disorder associated with decreased  
 CC expression or activity of NEUP. Antagonists of NEUP are useful for  
 CC treating or preventing disorder associated with increased expression  
 CC or activity of NEUP. NEUP or their fragments or derivatives are  
 CC useful for treating neurological disorder such as epilepsy, ischemic  
 CC cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's  
 CC disease, Pick's disease, Huntington's disease, dementia and  
 CC Parkinson's disease. NEUPs are also useful for treating other  
 CC demyelinating diseases, bacterial and viral meningitis, prion  
 CC diseases including Kuru, Creutzfeldt-Jakob disease, nutritional and  
 CC metabolic diseases of the nervous system, neurofibromatosis, other  
 CC developmental disorders of the central nervous system, cerebral  
 CC palsy, neuroskeletal disorders, autonomic nervous system disorders,  
 CC cranial nerve disorders, spinal cord diseases, muscular dystrophy and

CC other neuromuscular disorders, peripheral nervous system disorders,  
 CC inherited, metabolic, endocrine, and toxic myopathies, mental  
 CC disorders including mood, anxiety and schizophrenic disorders, a cell  
 CC proliferative disorder such as actinic keratosis, arteriosclerosis,  
 CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective  
 CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal  
 CC haemoglobinuria, cancers of the adrenal gland, bladder, bone,  
 CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory  
 CC disorder such as acquired immunodeficiency syndrome (AIDS), Addison's  
 CC disease, adult respiratory distress syndrome, allergies, ankylosing  
 CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,  
 CC complications of cancer, hemodialysis, and extracorporeal circulation,  
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic  
 CC infections, and trauma. This protein was given the Incey ID no.  
 CC 3201753CD1.  
 CC  
 XX Sequence 856 AA;  
 SQ

Query Match 5.1%; Score 355.5; DB 21; Length 856;  
 Best Local Similarity 26.0%; Pred. No. 2,1e-16;  
 Matches 132; Conservative 69; Mismatches 187; Indels 119; Gaps 17;

QY 247 NSRVEPVGHADTGLEHINFSLDNWKLYEVNDGPGIGIHVVPFSGRGRLGLVRL 306  
 DB 245 nkeptvttsdaavdlstfk---nvghlelpkdgqglgla!-----seedt!sgv!k! 297  
 QY 307 EKGGAHEHNLFRENDCIIVRINDGDLNRNRFEOAOHMFROAMRPTIWFH-----VVP 359  
 DB 298 tchgvaaatdgrlkvgdqlavddelivgyprffs!lktakmtvkl!lihaenpdsqavp 357  
 QY 360 AANKQEYEOLOSSEKNNYSSRFSPOSOYIDNRSVNSAGLHTVQAPRLNHPROIDS 419  
 DB 358 sa-----agaagekknssqslmwpqsg-----spepsint! 391  
 QY 420 RLPHSAHPGKRP---PSAPASAPQNVFSTVSSGYNTKKGKRLNIO!KKTGEGFSGIT 476  
 DB 392 r-----stspaf!asdpacp-----!lpgcet-----!t!skgrt!g!sl!v 430  
 QY 477 SRQVYIGGSAPLYVKNILPRGAIODGRLKAGDRLILEVNGVDLVGKSOEVSILRSTKM 536  
 DB 431 ggsdtll!ga---!l!heveyeegaackdgr!wagdq!levngid!l!r!katheda!nv!l!rq!cpq 488  
 QY 537 EGTVSLVFEQDAFHRRLNAPRSQMQIPKTKAEDDEDIVLPDGTREFLTFEVP!LND 596  
 DB 489 r---vrltlyrdeapy-----keeevcd-----!t!le!q!k! 517  
 QY 597 GSAGLGVSVKGNRSKENHADLGIFVKSIIINGGAASKDGLR!VNDOL!AVNGESILGKTKNO 656  
 DB 518 p!kg!sl!g!st!v!g!k!n---!dt!g!v!f!s!d!v!k!g!l!a!d!a!d!g!l!m!v!n!g!e!d!v!r!n!a!t!g!e 572  
 QY 657 DAMETLRRSMSTEGNKGMIQ---LIVARISKCNELSPSGSPPELPIPTALDDNERR 713  
 DB 573 aavaall!k!c!s!i!g!t!v!t!l!e!v!g!t!k!e!g!p!h!s!e!r!t!r!p!s!q!s!q!v---s!e!g!s!l!s!f!f!p!l!s!g!s!----- 624  
 QY 714 ISHLSVSGIEGIDESPRAALSRING 740  
 DB 625 -----stees!ess!sk!k!na!l!a!s!e!l!q! 645

RESULT 14  
 AAY53753  
 ID AAY53753 standard; Protein: 2037 AA.  
 XX AAY53753;  
 AC  
 XX  
 XX 22-FEB-2000 (first entry)  
 DE Amino acid sequence of the MMS2 protein.  
 XX  
 KW Human; MMS2; MMAC1; PDZ domain; tumour suppressor; tyrosine phosphatase;  
 XX scaffolding protein; cancer.

OS Homo sapiens.  
 PN W09958548-A1.  
 XX  
 XX 18-NOV-1999.  
 PD  
 XX  
 XX 07-MAY-1999; 99WO-US09969.  
 XX  
 XX 08-MAY-1998; 98US-0084740.  
 XX  
 XX (MRT-) MYRIAD GENETICS INC.  
 PA  
 XX  
 XX Bartel PL, Tavtigian SV;  
 PI  
 XX  
 XX WPI; 2000-053077/04.  
 DR N-PSDB; AA36453.  
 XX  
 XX  
 PT Nucleic acids and polypeptides representing human MMS2, useful for  
 PT detecting, diagnosing a predisposition to, and treating cancer -  
 PS Claim 1; Page 93-99; 112pp; English.

The present sequence represents human MMS2 protein. The MMAC1 protein binds to MMS2. The MMS2 protein has 11 post-synaptic density protein, disc-large, zo-1 (PDZ) domains and one or more of these domains interact specifically with the carboxyl terminal amino acids of MMAC1 (see AAY53753). Specifically, it appears that domain 7, 10 and 13 interact with MMAC1. Since MMS2 contains 11 PDZ domains and interacts with MMAC1, a known tumour suppressor having a region of homology with common biological pathways, MMS2 acts as a scaffolding protein in a interaction between MMAC1 and MMS2. It is believed that the activity of MMAC1. The MMS2 polypeptides, polynucleotides, fragments and or a predisposition to cancer and screening for agents that may be used to treat MMS2 and/or MMAC1 related cancer. The polypeptides and polynucleotides may also be used to treat cancer.

Sequence 2037 AA;

Query Match 5.1%; Score 355.5; DB 21; Length 2037;  
 Best Local Similarity 26.0%; Pred. No. 7.6e-16;  
 Matches 132; Conservative 69; Mismatches 187; Indels 119; Gaps 17;

QY 247 NSRVEPVGHADTGLEHINFSLDNWKLYEVNDGPGIGIHVVPFSGRGRLGLVRL 306  
 DB 1426 nkeptvttsdaavdlstfk---nvghlelpkdgqglgla!-----seedt!sgv!k! 1478  
 QY 307 EKGGAHEHNLFRENDCIIVRINDGDLNRNRFEOAOHMFROAMRPTIWFH-----VVP 359  
 DB 1479 tchgvaaatdgrlkvgdqlavddelivgyprffs!lktakmtvkl!lihaenpdsqavp 1538  
 QY 360 AANKQEYEOLOSSEKNNYSSRFSPOSOYIDNRSVNSAGLHTVQAPRLNHPROIDS 419  
 DB 1539 sa-----agaagekknssqslmwpqsg-----spepsint! 1572  
 QY 420 RLPHSAHPGKRP---PSAPASAPQNVFSTVSSGYNTKKGKRLNIO!KKTGEGFSGIT 476  
 DB 1573 r-----stspaf!asdpacp-----!lpgcet-----!t!skgrt!g!sl!v 1611  
 QY 477 SRQVYIGGSAPLYVKNILPRGAIODGRLKAGDRLILEVNGVDLVGKSOEVSILRSTKM 536  
 DB 1612 ggsdtll!ga---!l!heveyeegaackdgr!wagdq!levngid!l!r!katheda!nv!l!rq!cpq 1669  
 QY 537 EGTVSLVFEQDAFHRRLNAPRSQMQIPKTKAEDDEDIVLPDGTREFLTFEVP!LND 596  
 DB 1670 r---vrltlyrdeapy-----keeevcd-----!t!le!q!k! 1698  
 QY 597 GSAGLGVSVKGNRSKENHADLGIFVKSIIINGGAASKDGLR!VNDOL!AVNGESILGKTKNO 656  
 DB 1699 p!kg!sl!g!st!v!g!k!n---!dt!g!v!f!s!d!v!k!g!l!a!d!a!d!g!l!m!v!n!g!e!d!v!r!n!a!t!g!e 1753

QY 657 DAMETLRSMSTEGNKRGMIO---LIVARRISCKNELKSPGPPPELPETALDDRER 713  
DB 1754 aavaalkcslgtvltlevgrtkagpfhserpsqsgv-segslsftfpdisg----- 1805  
QY 714 ISHLYSGIEGLDESPSRNALSRIMG 740  
DB 1806 -----stseslesskhnalaseig 1826

RESULT 15  
AAV04733  
ID AAV04733 standard: Protein; 2070 AA.  
AC AAV04733;  
XX  
XX  
DT 06-JUL-1999 (first entry)  
XX  
DE Protein containing PDZ domain from clone 38-2-1b.  
XX  
XX PDZ domain; gene expression; human umbilical vascular endothelial cell;  
KW HUVEC; stimulation; tumour necrosis factor; TNF; protein binding;  
KW cell; proliferation disorder; cancer.  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO9907846-A1.  
XX  
PD 18-FEB-1999.  
XX  
PF 12-AUG-1998; 98WO-JP03603.  
XX  
PR 19-JUN-1998; 98JP-0189944.  
XX PR 12-AUG-1997; 97JP-0230356.  
XX  
PA (CHUGAI RES INST MOLECULAR MEDICINE INC.  
XX  
XX Funahashi S, Miyata S;  
XX  
DR WPI: 1999-167423/14.  
DR N-PSDB: AAX29910.  
XX  
XX  
PT Protein containing PDZ domain, whose expression is enhanced by TNF  
PT stimulation - plays an important role in protein/protein  
PT interactions and is used for screening for proteins for use in  
PT treatment of cell proliferation disorders such as cancer  
XX  
XX  
PS Claim 1: Page 167-176; 240pp; Japanese.  
XX  
CC This sequence represents a new protein containing a PDZ domain encoded  
CC by the clone 38-2-1b, whose expression in human umbilical vascular  
CC endothelial cells (HUVEC) is enhanced by stimulation with tumour necrosis  
CC factor (TNF) alpha. The new protein is used to identify proteins which  
CC bind to it (particularly to the PDZ domains) and the genes encoding them,  
CC for use in the treatment of cell proliferation disorders such as cancer.  
XX  
SQ Sequence 2070 AA:

Query Match 5.1%; Score 355.5; DB 20; Length 2070;  
Best Local Similarity 22.5%; Pred. No. 7,8e-16;  
Matches 188; Conservative 122; Mismatches 302; Indels 225; Gaps 37;

QY 230 GKWLKQEDDGDGTEDNSRVEPVGHADTGLEHHPNFSLIDMVKLYEVPNDGPGIIGHV 289  
DB 1119 grdipelpereeg-egeseelqntays-----nwnqprivelwrepeks--lgslv 1167  
QY 290 PPSARCGRL-----GLVVRLEKGGAEHBNLFRENDCTVRINDGILRNRRFE 338  
DB 1168 -----ggrgmgarlsngewmrgifikhvledrpaqngtlkpgdrivewgmdlrdashe 1222  
QY 339 QAOHMFROAMRPIIFHFVHPANKEOYEQLSQSEKNKNYSSRFPDSOYIDNRSVNSAG 398  
DB 1223 qaveairka-qnrv-fmwgslnrprksplpmlnlypkynfstnptadslqina-- 1278

QY 399 LHTVORAPRLNHPPEQIDSHSLRPLSHANPSGKPPSAPAPQNVSTTVSS----- 449  
DB 1279 -----dkapsqseese-----pekaplcsvpppppsafemsgshltgssaeaklsqvd 1325  
QY 450 -----GYNFKKIGKRLN-----IQLKKTGCLGSGSIT-SRQVYTGSGAPRYVKILP 495  
DB 1326 kedefgyswknirerygtltgelhmlelekyhsglglstagnkd--rsmvflvygldp 1382  
QY 496 RGAALDGRKAGDRLELVNGVDLVGKSGEEVSLSTKMEGTVSLVPRQEDAFPRE 555  
DB 1383 ngaagkgdgrlqiadelleiingqllygrshgnassllkcapsk--vkiffrnkdvngma 1440  
QY 556 L---NA-EPQOMQIPKETADEDDIVLTPDGTREFLTF-----EVLNDSGSAGLCVSV 605  
DB 1441 vcpnaveprpsnsenlnqketepvtlsdaavdlasfkvnqhulp-kdgqglglals 1499  
QY 606 KGNRSKENHADLGIFVASTINGCAKADGRLRVNDQILVANGESLIG-----KT- 654  
DB 1500 edtlis-----gvllkslshghvaatdgrlkyvgdqlavdvdeivgyplekflsllkta 1552  
QY 655 -----NDDAMETLRSMSTEGNKRGMIOIIVA-----BRISKCNELKS 692  
DB 1553 kmrvklitihaeinpqsqavpsaagaasgekknssqslmvpqsgspesistnrtstrcpa 1612  
QY 693 PGSPPG--PELP-ETALDDRER-----ISHSLYSGIEGLDESPSR 731  
DB 1613 faadpatcpllpccettelskgrtglslvgstdllgafllheve--egaackdgr 1670  
QY 732 NAALSRIKESG-----KYQLSPYVM---PQDDTVIIEDRLPLVPHLSIQ----- 776  
DB 1671 lwaqdqllvevngldlrkathdeainvrltqprvrltlyrdeapgykeeevcdltlelqk 1730  
QY 777 -----SSSSSHDDVGFVTADAGTWAKAASPSADCSLSP-----VDP 814  
DB 1731 kpgkglglslvgkrndtgvtrvsd---vkgyiadpdrlllgdqlillvngedvrvnasqea 1787  
QY 815 VLAFQREGFGROSN-----SEKR---TKQFSDAQULDFV-----KTRRSK 851  
DB 1788 vaallkcsllgtvltlevgrlkapgfserpsqtsqvsagslsftfpdlsstsesless 1847  
QY 852 SMLGTDENK-LNTVDDOKAGSPSRDVPGLKSSLSLEQTAFAEVTIANDIP 907  
DB 1848 skhnalaseigrltvemkk--gpldsiglsiaqvgvsl-----gdvp 1889

Search completed: July 24, 2002, 14:29:50  
Job time: 11382 sec







```

QY 597 GSAGLVSVKGNKRNKHNADLGIFFKSIINGAASKOGLRVNDQLIANGESILGKTNO 656
Db 1699 PGKLGISIVGKRR-----DTGTFVSDIYKGIADADRLKMGDOILWANGEDVRNNTQ 1753
QY 657 DAMETLRRSMSTGKNGKMIQ---LIVARRISKCNELKSPSPGPPELPIETALDDERR 713
Db 1754 AVAALAKCSIGVITTELVGKIKAGPFPHSRRPSOSQV--SESLISFTFPLSGS----- 1805
QY 714 ISHSILYSGIEGLDESPSRNAALSRIING 740
Db 1806 -----STSESLSSSKKNALASLEIIG 1826

RESULT 2
US-09-233-086-3
; Sequence 3, Application US/09233086
; Patent No. 6337192
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Taviglian, Sean V.
; TITLE OF INVENTION: MMSCI - An MMSCI Interacting Protein
; FILE REFERENCE: MMSCI - An MMSCI Interacting Protein
; CURRENT APPLICATION NUMBER: US/09/233,086
; EARLIER FILING DATE: 1999-01-19
; EARLIER APPLICATION NUMBER: US 60/071,861
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-233-086-3

```

```

Query Match 4.8%; Score 340.5; DB 4; Length 1881;
Best Local Similarity 21.0%; Pred. No. 2.3e-16;
Matches 185; Conservative 102; Mismatches 266; Indels 327; Gaps 35;

QY 124 ELEVPVSLRANMPLHVRSSDPALIGLSTVSDFSSE-----EPRSKNTMTSTA- 177
Db 877 ELXODPSMELVPLSHIQEATPV-----PSVNEHFGTOWLHNDPESQBARGRITVY 931
QY 178 -----GFLKONTAGSPKTCRRKD-----ENTRSLPRDTSNMSNOFQ 214
Db 932 SOEAQPYGCPENVMKENVFMELEPSVSTEGNSOQGRFDLEINISLAK-TSLDLMIP 990
QY 215 RDNARSLASHPMGKMLEKQDQEDGTEEDNSRVEPVGHADTGL----- 260
Db 991 NDVQGPSLIDLPVQAQ--RREQEDLPLYOHQATRVISKASATYGMLSKRAITDTCELP 1048
QY 261 -----EHPNPSLDMVNLVEY---PNDGGPLGIHVVPFARGRTL-----G 300
Db 1049 REEGEGEETPNFSGMPPIVEIFREPNVS--LGISIV-----GGQTVIRLNGEELG 1101
QY 301 LTVKRLKGGKREHNLRENDCIYRINDGLNRREFEOAHMFROAMRTPIWF----- 355
Db 1102 IRIKOVLEDSRAGKTNKATGDKITLEVSGVDLQASHSEAVEAIKNA-GNPVVFIVQSUS 1160
QY 356 ---HYVP-----AANKEQ-----YQLSQSEKNNYSSR 381
Db 1161 STPRVIVPNHNRANKITSNODTQEKKEKROGTAPPPKMLPPPYKALDDSDENEEDA 1220
QY 382 FSPD---SQYID-----NRVNSAGLH----- 400
Db 1221 FTDOKINGRADLPGEHITELKDKNGLGLSLAGNKRSRMSIFVYVINGPRAADCR 1280
QY 401 -----TVORAPRLNHP 412
Db 1281 MHGDELEITNNOILYGRSHQNASALITKAPSKVYLVIIRNEDAVNMAVYPPFVPSSP 1340
QY 413 EQIDSHS-----RLHHS-----AHPGSKPPSPAP 435

```

```

Db 1341 SSIEDQSTPISEEDGSLVGIKQLPESESEKFLAVSQMKQOKPRTVSSQDITLAP 1400
QY 436 ASAPQNV-----FSTVSSGYNTMKI--GKRLNQLKKGITGLCFSTSDVYIGG 484
Db 1401 ASSYSTADPFTGYGQAPLSDPATCPVPGQEMTIEISGRGGLSLI-----VGC 1454
QY 485 S---APIYVKNLIPRGAALODGRKAGDRLLLEVGVDLVKSQSEVYSLRSTKMGTV 540
Db 1455 KQPLNAIVIHVEYEGGAARGLMAGDQILLEVGVDLRNSSHEAITALROTPOK--V 1512
QY 541 SLVFRQEDAFHPRRLNAEPQOMQIPKTKADEDIYLTPTGTEFLTFEYPLNDGSGAG 600
Db 1513 RLAVYRDE-AHYRDEENLE-----IFPVDLQKRAKRG 1543
QY 601 LGVSVKGNKRNKHNADLGIFFKSIINGAASKOGLRVNDQLIANGESILGKTNDAME 660
Db 1544 LGISIVGKRRNS-----GVFTSDIYKGAADLDRLTQGOILVNGED-MRNASGEIYA 1597
QY 661 TLRRSMSTGKNGKMIQILVAR-----RISKCNELKSPSPGPPE-----LPETAL 707
Db 1598 TLKLC-----AQGLVLEIGRLRAGSMTSARTTSQNSQSOASHSCHSPFAFVITGL 1651
QY 708 DD--RERRISHSLYSGIEGLDESPSR--NAALSRIINGES 742
Db 1652 QNLVGTAKRVSDPSQKN-SGTDMEPRIVEINRELSDMLGIS 1690

```

```

RESULT 3
US-09-080-855-12
; Sequence 12, Application US/09080855A
; Patent No. 6083721
; GENERAL INFORMATION:
; APPLICANT: Saras, Jan
; APPLICANT: Franz, Petra
; APPLICANT: Aspenstun, Pontus
; APPLICANT: Helman, Ulf
; APPLICANT: Goner, Leonel Jorge
; TITLE OF INVENTION: PARC, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPBL
; FILE REFERENCE: L0461/7030
; CURRENT APPLICATION NUMBER: US/09/080,855A
; EARLIER FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: 08/805,583
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-080-855-12

```

```

Query Match 4.7%; Score 332; DB 3; Length 2466;
Best Local Similarity 20.9%; Pred. No. 1.5e-15;
Matches 193; Conservative 133; Mismatches 353; Indels 244; Gaps 32;

QY 80 DEODPHGGDGTSSASTGQSPSEIFGSELG-----TNNVSAFOPYQATSEI--E 126
Db 879 DADDI-----ERASFRSLMQLQASVAGFMGGRALISTGSLASTLNKLAVRPLVSQAEILKR 934
QY 127 VTPSVLRANMPLH--VRSSDPAL-----IGLSTVSDFSSEEPSRRNPTRWSTTAGEF 179
Db 935 LSCSPLSLYQPLQNSKKEKNDKASWEKPREMKSYHDLQSASLYHHRKN----- 984
QY 180 IKONTAGSPKTCRRKDENYRSLPR--DTSNMSNOFORDNARSLASHPMGKMLEKQEQ 238
Db 985 VIVNNEPPTQVETALVGKPSHOMKSRDASLQAGVTVLNKSKSVASIN-----RSP 1034
QY 239 DEDGTEEDNSRVEPVGHADTGLHHPNPSLDM-----VLYVEVPNDG--PL 284
Db 1035 ERKHESDSSSIDPPQA-----YVLDVLRKRWSTVSSPERETLVNMLKRDARYGL 1085

```





```

APPLICANT: HELDIN, CARL-HENRIK
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSER: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,804
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/596,291
FILING DATE: 09-AUG-1996
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: LC461/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2465 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-100-804-3

```

Query Match 4.7%; Score 327.5; DB 3; Length 2465;  
Best Local Similarity 20.9%; Pred. No. 3.2e-15;  
Matches 194; Conservative 134; Mismatches 346; Indels 253; Gaps 34;

```

80 DEODPHHGDDGTASSTGTQSPETFGSELG-----TNNVSAFOPYQATSET--E 126
Db 879 DAODI-----ERASFRSLNLAESVGRFNMGRALSTGSLASTLKLAVRPLSVQAEILKR 934
QY 127 VTPSVLRANPLH--VRSSDPAL-----IGLSTVSDSNFSESEPEPRKKNPTKSTTAGF 179
Db 935 LSCSELSTLYOPLQNSSKEKNKDKASWEKPREMSKSYHDLQASLTPYPRKN----- 984
QY 180 LKONTAGSPKTDCKRDKDEYRSLPR-DTSNWSNQFORNARSSLSLSPVGVKMLEKOEQ 238
Db 985 VLVNMRPPQTYAELVGKFSHOMSRSDASLACVTKLNNKSVASLN-----RSP 1034
QY 239 DEDGTEEDNSRVEPVGHADTGLEHINFSLDDM-----VKLVEVPNDGG-PL 284
Db 1035 EKRKHSSDSSSTIEDPQA-----YVLVDLHKRMSTVSSPERETITLVNKKDKAKYGL 1085
QY 285 GHVVPFSARGR---TGLGLVRLKLGKGAHEMLFRENDCIVRINDGLRNRRPEQA 340
Db 1086 GQDIT-----GGEKMETDGLGIFISSVAPGPGPADFHGCLAKFGDRLISVNSYLEGVSHHA 1140
QY 341 OHMFROA-----MRTPLIMFHY-----VPRANKEOY- 366
Db 1141 IETLQNAPEDDVLTVISQPEKISKVSPSTVYHLNEMKNTWKKSYSMODSAIDSSSKDHW 1200

```

```

QY 367 -----EQLSQSEKNMYSSRF----- 382
Db 1201 SRGTLRHISNSRPGSGGLREGSLSSQDSRTESLSQSYVNGFPAHLDDQFTWQESOHG 1260
QY 383 SPDSQYIDNRSVNSAGLHTVQ---RAPRLNHPPEQID-SHRLPLHSPGKPPSPAPASA 438
Db 1261 SPSPSVYSKATKEKETETFDNSQNSKTKKPGISDVVDYDSRGSGSDMEAAVSSSDHQTPKOE 1320
QY 439 PQNVSTFVSSGYNT-----KKIGKRLNIQLKGTBELGFSIT---SRDVTIGGSARITV 490
Db 1321 SSSSVNTSNKMNFKTESSSPKPGDIFEVELAKNDSGLISTVGCVNTSVRHG---ITV 1377
QY 491 KNILPRGAALODGRILKAGDRLIEVNGVDLVGKSOEEVYSLRSTKMGVTSILVFRQEDA 550
Db 1378 KDVIPOGAASDGRIRIHGDDVLAVNGSLGATHKQAVETLRNT---GVYVHLLLEGGQS 1434
QY 551 FHPRE-----LNAEPSOMQIP---KEYAEDEDIVLTPDGTREFTFEVPL-ND 596
Db 1435 PTKEHVPVTPQCTLSDONAGOGPEKVKTKTQYKDYSPV-----TEEMTFEVLKFNKS 1488
QY 597 GSAGLGYSVYKNSKEHADLGIYFKSIINGAASKDGRLVNDQLAVNGESILGKTQ 656
Db 1489 SGLGFSEFSREDNLIPEDINASIVAKKLFAGQPAAESGKIDVGDVILKVGASLGLSQ 1548
QY 657 DAMETLRRSMSTEGNKGMIQLIYARRISKCNELKSPGSPPELPJETALDDERRISH 716
Db 1549 EVISALRGTAPE-----VFLLCR-----PPGVLPETDNL-----LT 1582
QY 717 SLVSGIEGLDESPSRNALSLRI-----MGESGKYOL-SPTVNMPODDVILIEDR 765
Db 1583 PLOSPAGVLTPLNSSKSDQSPCEQSTSDDENEMSKKCKOCKSPSRDYSVSDSGSGEDD 1642
QY 766 LPVLPRLHSDQS-SSSHDDVGFYTLADAGTAKAATIS-----D 802
Db 1643 LVTPAPANTISVTSWSSALHQTLSNNVSOAQSHHEAPKQEDPTICTMEYYPOKIPKPEFED 1702
QY 803 SADCSLSPDVPVLAFOREGFGROSMS 829
Db 1703 SNPSPLPDDMAPQSGYQPOSEASASSS 1729

```

RESULT 7  
US-09-290-640-46  
; Sequence 46; Application US/09290640  
; Patent No. 6204055  
; GENERAL INFORMATION:  
; APPLICANT: Dean, Nicholas M.  
; APPLICANT: Marcussen, Eric G.  
; TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling  
; FILE REFERENCE: ISPH-0351  
; CURRENT APPLICATION NUMBER: US/09/290,640  
; CURRENT FILING DATE: 1999-04-12  
; NUMBER OF SEQ ID NOS: 85  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 2485  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-290-640-46

Query Match 4.6%; Score 324.5; DB 4; Length 2485;  
Best Local Similarity 20.7%; Pred. No. 5.4e-15;  
Matches 193; Conservative 134; Mismatches 361; Indels 245; Gaps 32;

```

80 DEODPHHGDDGTASSTGTQSPETFGSELG-----TNNVSAFOPYQATSET--E 126
Db 879 DAODI-----ERASFRSLNLAESVGRFNMGRALSTGSLASTLKLAVRPLSVQAEILKR 934
QY 127 VTPSVLRANPLH--VRSSDPAL-----IGLSTVSDSNFSESEPEPRKKNPTKSTTAGF 179
Db 935 LSCSELSTLYOPLQNSSKEKNKDKASWEKPREMSKSYHDLQASLTPYPRKN----- 984

```

```

QY 180 LKONTAGSPKTCRRKDENYSLPR-DTSMNSNOFORDNARSSLSASHPWKMLEKQO 238
Db 985 VYINMEPPQVIAELVGRPSHOMSRSDAESLAGVTKLNNSKSVASLN-----RSP 1034
QY 239 DEDGTEENSRVEYGAH-----DTGLEHIFPESLDDM-----VKLV 275
Db 1035 ERKKESSSSSIEDPGQVYVLGKMTMHSSGNSSQVPLKENDVILHKRWSIVSSPEREITLV 1094
QY 276 EYPNDGG-PIGIHVVPFSAAGRTILGLVLRLEKKGAKAHENILFRENDCIYIINDGDIRN 334
Db 1095 NLKMDAKYGLGFOIIGGEKKGRIDLGIPTISSVAPGPPALDCCLPGRILISVNSVLEGG 1154
QY 335 RRFEOAHMFROA-----MRTPIWFHV-----VPA 361
Db 1155 VSHHAILEILQNAPEVTVLVIQPKRKISKVPSTPVHLTNMKMYMKSSMODSAIDSS 1214
QY 362 NKREY-----EOLSOSEKNNTYSSRF-----382
Db 1215 SKDHHSKGTLRHISENSFGPSGLREGSLSSQDSRTESLSLSOSQVNGFASHLDQTM 1274
QY 383 -----SPDSOYIDNRSVNSAGLHTVQ--RAPRLNHPPOID-SHRLPHSAHPGKRP 432
Db 1275 QESOHGSPSPVYSKATEKEFTTDSNOSKTKKPGISIDVTYSDRGSDMDPATYSSODH 1334
QY 433 SAPASAPQNVSTYVSSGYNT-----KKIGKRLNIOLKKGTEGFSIT--SRDVTIGG 484
Db 1335 QTPKQESSSVNTSKMNFKTFSSPPKPGDIFEVELAKNDNSLGIYVGGVNTSVRHGG 1394
QY 485 SAPYIVKNILPRGAIIODGRKAGDRLEIVNGVDLVKGOEVSILSTKMEGTVALV 544
Db 1395 ---IYVAIVIPOGAAESDGRHKGDVILAVNGVSLGATHKQAVETLNT--GOVYVHL 1448
QY 545 FRODAFHPRE-----LNAEPSOMQIP--KETKAEEDIVLPDSTREPLTFEY 591
Db 1449 LEKGQSPTEKHEHPVTPQCTLSDONAOGGPEKVKTTQVADYSFY-----TEENFEY 1502
QY 592 PL-NDGSAGLVSVKGNRSKENHADLGIYFKSIINGAASKDGRLEVNOLDIYVNGEST 650
Db 1503 KLRKNSGGLGFSPSRDNILPEQINASTIVRYKKLPPOCPAAESGKIDVGIILKVGASL 1562
QY 651 LGKTNODAMETLRRSMSTEGNKGRIQILTVARISKCNELKSPSPGPELPTERLADR 710
Db 1563 KGISOOEIVSALRGTAPE-----VFLLCLR-----PPGVLPETIDRL-- 1600
QY 711 ERRIHSILXSGIEGLDPSPRNALISRI-----MGESKTYOL-SPTVNMPODDTV 759
Db 1601 ---LTPLOSPAQVLPNSSKSDSQPCVEGQSTSSDENKSDSKCKSPSRKDSYSDSS 1656
QY 760 IIEDRLPVLPRPHLSDOS-SSSSHDDVGYVTADAGTMAKAALIS-----801
Db 1657 GSGEDDILVTAPANISNSTGSSALHQTLSNMVSOAQOSHHEAPKSDHTICTMFFYYPKAI 1716
QY 802 ---DSADCSLSPDVPYLAQREGFGROGMS 829
Db 1717 KPFEEDSNPSPLPPDMAPOGYOQSESASSSS 1749

```

```

RESULT 8
US-08-410-804-1
; Sequence 1, Application US/08410804
; Patent No. 5632994
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Sato, Yukaaki
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive, Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States

```

```

ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/410,804
FILING DATE: 27-MAR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/259,514
FILING DATE: 14-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn
REGISTRATION NUMBER: 31,815
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 610 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-410-804-1

```

```

Query Match 3.9%; Score 275.5; DB 1; Length 610;
Best Local Similarity 24.8%; Pred No. 2,7e-12;
Matches 132; Conservative 77; Mismatches 193; Indels 131; Gaps 19;
QY 360 ANKEQYEQLOSSEKN-----NYSSRFSPDSOYIDNRSVNSAGLHTVQAPRLNHP 412
Db 12 ATEKEFTTDSNOSKTKKPGISIDVTYSDRGSD--MDEATYSSQDH-----QTP 59
QY 413 EOLDSRLPHSAHPGKRPASAPQNVSTYVSSGYNTKKIGKRLNIOLKKGTEGIL 472
Db 60 KO-ESSSVNTSKMNFKTFSSPPKPGDIF-----EVELAKNDNSLG 101
QY 473 FSIY-----SRDVTIGGSAPIYVKNILPRGAIIODGRKAGDRLEIVNGVDLVKGSQ 524
Db 102 ISVTILFPGKGVNTSVRHGG---IYVAIVIPOGAAESDGRHKGDVILAVNGVSLGATH 158
QY 525 EEVNVILSTMEGTVALVFRQDAFHPRE-----LNAEPSOMQIP--KETKA 571
Db 159 KOAVETLNT--GOVYVHLLEKGQSPTEKHEHPVTPQCTLSDONAOGGPEKVKTTQV 215
QY 572 EDEDIVLPDGTREFTLFEVPL-NDGSAGLVSVKGNRSKENHADLGIYFKSIINGAA 630
Db 216 KDYSFV-----TEENFEYKLEKNSGGLGFSPSRDNILPEQINASTIVRYKKLPPOCPA 269
QY 631 SKDRLRVNOLDIYVNGESTLKTNODAMETLRRSMSTEGNKGRIQILTVARISKCNEL 690
Db 270 AEGKIDVGIILKVGASLKSQOEIVSALRGTAPE-----VFLLCLR----- 314
QY 691 KSPGSPGPELPTERLADRERRISHLSYSGIEGLDPSPRNALISRI-----MG 740
Db 315 ---PPGVLPETIDRL-----LTPLOSPAQVLPNSSKSDSQPCVQPSDENKMS 363
QY 741 ESGKYOL-SPTVNMPODDTVIIEDRLPVLPRPHLSDOS-SSSSHDDVGYVTADAGTMAKA 798
Db 364 DKSKQCKSPSRKDSYSDSSGSGEDDILVTAPANISNSTGSSALHQTLSNMVSOAQOSHHEA 423
QY 799 AIS-----DSADCSLSPDVPYLAQREGFGROGMS 829
Db 424 PKSDHTICTMFFYYPKAIKPKPFEEDSNPSPLPPDMAPOGYOQSESASSSS 476

```

```

RESULT 9
US-08-259-514-1
; Sequence 1, Application US/08259514

```

Patent No. 5747245  
GENERAL INFORMATION:  
APPLICANT: Reed, John C.  
APPLICANT: Sato, Takaaki  
TITLE OF INVENTION: FAS ASSOCIATED PROTEINS  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cathryn Campbell  
STREET: 4370 La Jolla Village Drive, Ste 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/259,514  
FILING DATE: 14-JUN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 9954  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 610 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-259-514-1

Query Match 3.9%; Score 275.5; DB 1; Length 610;  
Best Local Similarity 24.8%; Pred. No. 2.7e-12;  
Matches 132; Conservative 77; Mismatches 193; Indels 131; Gaps 19;

QY 360 AANKEQYEQLOSSEKN-----NYSSRFPDSQYIDNRSVNSAGLHTYQAPRLNHPP 412  
DB 12 ATEKEFTDSDNSQSKTKPGISDVTDYSDRGDSD--MDEATYSSQDH-----QTP 59  
QY 413 EQDISHRLPHSAHPGKPPSPAPASAPQVNFSTVSSGYTKIKIGRLNIQLKKGTEGLG 472  
DB 60 KO-ESSSVNTSKMFKTPSSPPKPGDIF-----EVELAKNDNSLG 101  
QY 473 FSIT-----SRDYTIGGSAPITYVKNILPRGAIIODGRKAGDRILEVNGVDLVKRSQ 524  
DB 102 ISVTVLFDKGVNTSVRHG---IYKAVIPQGAESDGRHKGDRVLAVNGVSLGATH 158  
QY 525 EEVYSLRSTKMEGTSLVLFRODAFHPR-----LNAEPSOMQIP---KETKA 571  
DB 159 KQAVETLRNT---GQVYHLLLEKGSPTSKEHVPVTPQCTLSQDNAGOGPEYKKTQY 215  
QY 572 EDEDIYVLPDGTRELFTEFVPL-NDSGSAGLGVSVKGNRKENHADLGIYFKSLINGCA 630  
DB 216 KDYSFY-----TEENTFEYKLFKNSGGLGFSRSREDNLIPEDINASIYVKKLFPQPA 269  
QY 631 SKDRLRVNDOLIAVNGESILKTNODAMETLRRSMSTEGNKKGMQLIYARRISKNEL 690  
DB 270 AESGKIDVGVILKVGASLKLGSQCEVISALNGTAP-----VFLILCR----- 314  
QY 691 KSGSPGPEPLPTETALDRERRISLSLISGLGDESPRNALSTI-----MG 740  
DB 315 ---PPGVLEIDTAL-----LTPLOSAPQVLPNSSSKDSSPCSCVEOSTSDENEMS 363  
QY 741 ESGKYOL-SPTVMPPDDVTIIEDDLRPLVLRPHLSDOS-SSSSHDVGVGTADAGTWAK 798  
DB 364 DKSKKCKSPSRSDYSTDSGSGEDDLVTAPANISNTWSSALHQTLSNWNVSQAQSHHEA 423

QY 799 AIS-----DSADCSLSPDVPVLAFQREGGRQSMS 829  
DB 424 PKSQEDTICMTEYYPQIKIPKREFEDSNPSPLPDMAPGSGYQPOSESASSS 476

RESULT 10  
US-08-858-311-1  
Sequence 1, Application US/08858311  
Patent No. 5876939

GENERAL INFORMATION:  
APPLICANT: Reed, John C.  
APPLICANT: Sato, Takaaki  
TITLE OF INVENTION: FAS ASSOCIATED PROTEINS  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cathryn Campbell  
STREET: 4370 La Jolla Village Drive, Ste 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/858,311  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/410,804  
FILING DATE: 27-MAR-1995  
APPLICATION NUMBER: US 08/259,514  
FILING DATE: 14-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 1389  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 610 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-858-311-1

Query Match 3.9%; Score 275.5; DB 2; Length 610;  
Best Local Similarity 24.8%; Pred. No. 2.7e-12;  
Matches 132; Conservative 77; Mismatches 193; Indels 131; Gaps 19;

QY 360 AANKEQYEQLOSSEKN-----NYSSRFPDSQYIDNRSVNSAGLHTYQAPRLNHPP 412  
DB 12 ATEKEFTDSDNSQSKTKPGISDVTDYSDRGDSD--MDEATYSSQDH-----QTP 59  
QY 413 EQDISHRLPHSAHPGKPPSPAPASAPQVNFSTVSSGYTKIKIGRLNIQLKKGTEGLG 472  
DB 60 KO-ESSSVNTSKMFKTPSSPPKPGDIF-----EVELAKNDNSLG 101  
QY 473 FSIT-----SRDYTIGGSAPITYVKNILPRGAIIODGRKAGDRILEVNGVDLVKRSQ 524  
DB 102 ISVTVLFDKGVNTSVRHG---IYKAVIPQGAESDGRHKGDRVLAVNGVSLGATH 158  
QY 525 EEVYSLRSTKMEGTSLVLFRODAFHPR-----LNAEPSOMQIP---KETKA 571  
DB 159 KQAVETLRNT---GQVYHLLLEKGSPTSKEHVPVTPQCTLSQDNAGOGPEYKKTQY 215  
QY 572 EDEDIYVLPDGTRELFTEFVPL-NDSGSAGLGVSVKGNRKENHADLGIYFKSLINGCA 630

```

Db 216 KQYFV-----TEENTREVKLFKNSSGLGFSFREDLILPEQJINASIYRKKLFPQOP 269
QY 631 SKDRIRVNDOLIVNGSILGKTNOAMETLRBSMSEKRGKIOILIVRRIKSKNEL 690
Db 270 AESGIDGVDVLKNGSLGQEVTSALRGAP-----VILLDR----- 314
QY 691 KSPGSPPELPETALDRERRISLSLYSGIEGDESPPRNALSR-----MG 740
Db 315 -----PPGVLEIDAL-----LTPAQPAQVLPNSSKSSQSPSCVEQSTSDENEMS 363
QY 741 ESGYQL-SPVNNPQDDTVIIEDRLPVLPPHSDOS-SSSHDDVCFVADAGTAKA 798
Db 364 DKSKKQCKSPSRSDYSDSGSDDLVTAPANISNSTWSALHQLTLSNMVSOAQSHHEA 423
QY 799 AIS-----DSADCSLSPVDVPLAFOREGFCRQSM 829
Db 424 PKSQEDTCTMYFYQKIPNKEFEEDSNPSPLPDMAPGQSQPOSESASSS 476

```

```

RESULT 11
US-08-545-860D-48
; Sequence 48, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESS: No. 6040140ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10930
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/327,392
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/320,559
; FILING DATE: 11-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,443
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,094
; FILING DATE: 30-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/888,839
; FILING DATE: 27-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,093
; FILING DATE: 11-DEC-1991
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1612 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-08-545-860D-48

```

```

Query Match 3.7%; Score 260; DB 3; Length 1612;
Best Local Similarity 20.5%; Pred. No. 1,7e-10;
Matches 297; Conservative 190; Mismatches 450; Indels 510; Gaps 77;

```

```

QY 49 HRLHGGGIIIDLDIICDV-----ADKDRLYAVFDEODPH----- 86
Db 307 HSDEGAKETI-LDDDECPLOIFREWPSDKGILVFLQKRPPDHPKTKKHLBEGTKPKG 365
QY 87 -GGDGTSSAST-----GTQS--PEIRG-----SELGT-----NNVSAF 116
Db 366 KERADSYGSGTLPPEKLPYLVELSPDQSDSRDKRYRLQSLSTVEVTEKIDMSIGL 425
QY 117 -----QPYA-----TSELYTPSVIRAMPLHVRSSPALI--GLSTVSQDSN-FSSE 164
Db 426 GPGIOPHHCIDLINMDGVVTVTPRSMADETVYEGORISTMTMQSMKQFGASHVFKFVD 485
QY 165 PSRKNP-TRWGTAGFLKONTAGSP-----KTCDBKRD-ENYRSLPRTSWMNSNOFODN 217
Db 486 PSQDHAKKRVYDGLMKVGRPHRGIVQETTFDGLDGHSGTALP--TSKSTRIDSDR 543
QY 218 ARSSLASH-----PMVG-----KMLEKQODEDGT-----EDNSRVEP 252
Db 544 VSSASTAERGVMKIRIVEGOQDVRQESRTQDASGPELLIPASIEFRESSESLSAI 603
QY 253 VGHADTGLEHT---PNFSLDMKLYEVPNDGPGILGHVYFVSARGGTLLILYKRLKG 309
Db 604 INTNNSSTVHKRLSPYVLYMACRYV-LSNOYRP--DISP-TEETHVIVAVNNMV--- 655
QY 310 GKAHENLFRENDCIV-----RINDGDLRRRFEQAOHMFQAMRPTIIMFHVPAANK 363
Db 656 ---SMMEGYIQOKNTAGALAFMMANASSELIN-FIKQDBLSRITLDADYLAHLVQMAFK 712
QY 364 ROYEQLOSSEKNNYSSRFPSPDSOYIDNRSVNSAG-----LHTVO-----RAPRLNH 410
Db 713 YLVHCL-QSELNNYMPA-----FLDDPEENSLOPRKIDYVLIHTLITAMSLIRRCRYN- 763
QY 411 PPEQIDSHRSLPHSAHPGKRPSPASAPAPVNESTVS---SGYTKKIGKRLNTQL--- 464
Db 764 AALTIQLFSQLPHFTN-----MMLEFNLVTDPPSGILCSHWGALIRQDLHI 810
QY 465 -----KKGTGEGFSITSRDV-----TIGSAPATYVNI-----LPGAIADGR 504
Db 811 EAMAKQGLLELAADCHLEIRIVQATLLTMDKYAPDDIDININSTCFKMSLQALQALQYVH 870
QY 505 LKAGRLIEVNGVDLVGKSQEEVVSLSRSTKMEGVSLIVROEDAFHPRELNAFSPQMO 564
Db 871 CAPDEPFP--TDLI-----ENVVVAVENTADE-----LARSQG--REVOLR- 908
QY 565 IPKTKARDEDIVL-----TPDGTREFLT----- 588
Db 909 -----EDPDQLPFLLPEDGYSQDVVRNIPNGIQEFLDPLQCGFRLLPHTRSPTGW 961
QY 589 ---FE-----VPLNDSGSAGLG--VSYKGNRSKENHADL 617
Db 962 TIYFEGADYESHLLARENTELQPLKKEPELITVTLKNGKNGLSIYAAKG---AGQKX 1017

```





```

Db 909 -----EDPDQLPFLPEDEGYSCDVYRNIPNGIEFLDPLQGRGFCRLIPHRSPGTW 961
QY 589 ---FE-----VPLNDSGSAGLG-VSKGNRSKENHADL 617
Db 962 TTFEGADVESHLEKRENTLEAOLRKEPELITVTLKKONGMGLSTVAAGK---AGDKL 1017
QY 618 GTFVKSIIINGAASKDGRLEAVNDOLLAANGEISLTKTQDMETLRSMST---EGNKG 674
Db 1018 GIYVSVKGGADVDGRLAAGDOLLSTVDGRSLVGLSERAELMTSTSSVITLVAAGK 1077
QY 675 -----MIOLIVARRISKCNELKSP-----GSPGPEL----- 701
Db 1078 AIYHGLATLLNOPSMMQRIISDRSGSGKPRKSEGFELYNNTQNSPSPQOLPAEYSE 1137
QY 702 PIETALDDR--ERRISHSLYSGIEGLDESPSRNALSRINGESGKQLSPTVNM----- 753
Db 1138 PKKLPPDDRLMKNRADHHSFPVANOPSPSGKSAVA--SGTTAKITSVGTGNCLEBOT 1195
QY 754 --PODDTVI-----EDDLPLVLP-----PHLSOSSSSSHD 784
Db 1196 PPPPEAVPIPTQYTYTTEYTFPPASKSODRM--APPQOMPVEYEKPRMHTDMSHS--- 1250
QY 785 VGFYTAAGTWAKAALISDSADCSLSPDVPLAFORGEFGQMSSEKRTQFSDASQLDF 844
Db 1251 -----IAIQRTVRSQELREDKATQL--ERRHLEA 1278
QY 845 VKTRKSKMDIGIADETKL--NTVDDQKAGSPSDVGPSTGLKKS-----SLESILQ 894
Db 1279 AMDRKSDS--DMWINGSSSLDSTSSOEHLNHSKSVTPASTLTKSGPGMKTPAIPATP 1337
QY 895 TAAVEYTLNGDIPFRPRRIIRGRGCSFRAAIDKSDKRAVDDDEGMEETLEEDTEE 954
Db 1338 VAVSQ--PIRTDLPPEPPPPV-----HYAGFDMSMDLLPPLPPS 1377
QY 955 SRSRGREVSSTAQPSHSLEROMNGNOEKGDKTRD--KKDTGKEKKDKDEKDKMA 1012
Db 1378 ANQIGLPSAOVA-----AERKKEEHORWYKEKAELEPERR----- 1416
QY 1013 KKGMLKGLDMFRGKHRRKDDIKETGKTIQ-----ESFTSEERIRIMQOERRIOA 1065
Db 1417 -----KRROQR--KLGOMKQSLNPAPFSPLAQO--MKPKPSTLO 1455
QY 1066 KTRB--FRROARERDYAEIODFHRTGCDDELMTGVSSEYSGMALNAP----- 1114
Db 1456 RPQETVIRELQPOO-----RTERRD--LOYITVSKESLSDSLSPDPMKRAKE 1506
QY 1115 --OSPEGHMDALYAQVKPRKSPSPVDNSRSTSNDRIOQLROEF-----GQAKQ 1166
Db 1507 KLEKQOQHIIYDMLSKETIOE--IOSKP-----DRSAEES--DRLKIMLEMOQKRLDESKO 1559
QY 1167 -DEVED 1172
Db 1560 KDEDDEE 1566

```

## RESULT 13

```

US-09-157-420-1
Sequence 1, Application US/09157420
Patent No. 6180760
GENERAL INFORMATION:
APPLICANT: TAKAI, Yoshimi
APPLICANT: NAKANISHI, Hiroyuki
APPLICANT: MANAI, Kenji
APPLICANT: WADA, Manabu
TITLE OF INVENTION: ACTIN FILAMENT-BINDING PROTEIN "L-AFADIN"
FILE REFERENCE: 98-10442/LC(PMC)/653
CURRENT APPLICATION NUMBER: US/09/157,420
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1829

```

```

; TYPE: RPT
; ORGANISM: rat
US-09-157-420-1

Query Match
Best Local Similarity 3.78; Score 260; DB 4; Length 1829;
Matches 179; Conservative 131; Mismatches 329; Indels 272; Gaps 35;

QY 530 ILRSTKMEGVYSLVLRQEDAFHPRLENAEPSONQIPRETKAEDIEDIVLPDQTRREFLTF 589
Db 974 LVPHTRSPGTWTIYF--EGADVESHLEKRENTLETOPLRKEPE-----VIT----- 1016
QY 590 EYPLNDGSAGLG-VSKGNRSKENHADLGIYKSIINGAASKDGRLEAVNDOLLAANGE 648
Db 1017 -VTLKRONQMGSLIYAAGK-----AGDKLGIYVSVKGGADVDGRLAAGDOLLSTVDGR 1071
QY 649 SLTGKTNQDAMETLRSMST--EGNKG-----MIOLIVARRISKCNEL 690
Db 1072 SLVGLSERAELMTSTSSVITLVAAGKAGIYHGLATLLNOPSMMQRIISDRSGSKPR 1131
QY 691 KSP-----GSPGPEL-----PIETALDDR--ERRISHSLYSGIEGLDESPS 730
Db 1132 KSEGEELYNNSAONGSPESPQWTEYSEPKKLPPDDRLMKNRADHHSFPVANOPSPSG 1191
QY 731 RNALSLINGESGKQLSPTVNM-----PODDTVI-----EDDR 765
Db 1192 GKSPYT--SGTTAKITSVGTGNCLEBOTPPPEAVPIPTQYTYTTEYTFPPASKSODRM 1249
QY 766 LPVL-----PHLSOSSSSSHDVGFTVADGATWAKAALISDSADCSLSPDVPLA 817
Db 1250 APVQOMPVEYEKPRMHTDMSHS-----IA 1275
QY 818 FOREGFGQMSSEKRTQFSDASQLDFVTKRKS--MDIGIADETKNTVDDQKAGSPSR 876
Db 1276 IQRTVRSQELREDKATQL--ERRHLEA-----SLESILQ 894
QY 877 DVGPSLGLKS-----SLESILQTAVA--EYTLNGDIPFRPRRIIRGRGCSFRAAI 929
Db 1335 SVTASLTITLTSGBRKMKTPAAVLPFPVAVSOPIRTDLPPEPPPP-----PHAY 1382
QY 930 DKSTDKPAVDDDEGMEETLEEDTESSRSGRSVSTAQPSHSLEROMNGOERKDTD 989
Db 1383 TSFDISMD-----LPLPEPPANOAAPOSNOVAAEERKKEEHORWYKEKA--- 1430
QY 990 RKDKTGKREKDRKEDKAKKARKMLKGLDMFRGKHRRKDDIKETGKTIQIESFTS 1049
Db 1431 -RLEERERKREDEKRIKQKRTQSLNPASFPLAQKPEKPSLOR-----P 1478
QY 1050 EERIRBMQEOERIOAKTRERROARREDDYAEIODFHRTGCDDELMTG--GVSYEGSK 1108
Db 1479 QETVIRELQPOO--QPTTERRDLO-----YTTIS-----KEILSSDLSLPPMKR 1523
QY 1109 ALNARPOSPEGHMDALYAQVKPRKSPSPVDNSRSTSNDRIOQLROEF-----Q 1162
Db 1524 DAREKLEKQOQHIIYDMLSKETIHELQN-----KGDRTAEEDDRKLTMLMWOQKRLQ 1576
QY 1163 QAKODVEDEDRRRTYSFPQWPNARPATOSGRHSYSVENOMOROEERESSQAOQOYS 1222
Db 1577 ESKDKEDDEDEED-----DDVDTMLIMORLAERARALOQDEERRQ 1618
QY 1223 SLPRQSRKAASVSODSEWQNSYSGEGFOSAKENPRYSYQSGRNQYLGHGFMARML 1282
Db 1619 QOLEMKRRR--VEDVROEED-----GRHOERKRD 1650
QY 1283 TOELLROBO--RRKEQOMKQPPSGPSNDSYKQVDP-----SYAPKGP 1327
Db 1651 AEERKROEEGYSLERARRRQ-----HEEAARLLEPEEGSLRNPLPDVPEPPSS 1703
QY 1328 FRODVPPSPSQ 1338
Db 1704 SASSAPPPPPQ 1714

```



CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/045,632  
 FILING DATE: 19-MAR-1998  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/041,016  
 FILING DATE: 19-MAR-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Corless, Peter F.  
 REGISTRATION NUMBER: 33,860  
 REFERENCE/DOCKET NUMBER: 48147/1699-CIP  
 TELEPHONE: 617-523-6440  
 TELEFAX: 617-523-6440  
 INFORMATION FOR SEQ ID NO: 32:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1061 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-045-632-32

Query Match Best Local Similarity 3.2%; Score 222.5; DB 3; Length 1061;  
 Matches 196; Conservative 134; Mismatches 362; Indels 335; Gaps 41;

QY	215	RDNA	SSLSASHPMVGKMLEK-QEODEDCTEDNSRVEPV-GHADTGLEHINPNSL----	268
DB	116	RGAH	DDRNKRSRVVITCRPGPDRETGKPGDRLSTVDGIRLTGTHAEMSLIKQC	175
QY	269	-DMV	ALVE-----VPNDGPLGIHVVPFSARGRTLG-----LAVKRL	306
DB	176	GOE	ATLITDYDVSANDSVATASGPL--LVEVAKTPGASLGLVLTYSVCCNKQIVIDKI	232
QY	307	EKG	GAHEHNTFRENDCTVRINDGDLRNRFEQA-----QHMFQAMR	349
DB	233	KSAS	IADRCALHYVDHILSDTSMECTLAETOPLGNTTDOVKLEILPHHQRLALK	292
QY	350	PLIF	HVVPANKQEQYQLSOSSEKNYSSRPSDQYIDNRSVNSAGLHVQARPLN	409
DB	293	GP	-----DHKKIQRSDRLPW-----DPMASQCSVHT-----NHH	323
QY	410	HPPE	IDSHSLPHSAHPGSKPPSAP-----ASAPQNVSTVS-----SGYNTKK	455
DB	324	HNPH	PP-HCRVPALGPPKALTPNSPPAMVSSSPYSMSAYSLSLNMGTLPRLSYTSP	382
QY	456	IGK	RLNIQLKKTGEGLSITSRDVTIG-----	484
DB	383	RGTM	RRRLKKKPKSSLSLASSVGLAGOVVHTETTEVVLADPYTGFIQLDGSVPAT	442
QY	485	----	SAPLYKANILPRGAALQDRLKAGDRLIEVNGVDLVGKSQEEVVSLLRSTMECTVS	541
DB	443	ETLS	SPPLSYIEADSPAERCGVLIQGRVMAINIGIPTEDSTFEANQLRDSITSKYT	502
QY	542	L	-----LVFRQDAFH-----PRELNAE-----PSOMQ-----IPKETA-----	571
DB	503	LEIE	PVASVITPSSGTFHVKLPPKKHVEYLGITISSPSRKPDPPLVYSIDIKKGSVAHRT	562
QY	572	----	ED-----EDIV-----LTPDGTRE-----FLTFEV	591
DB	563	GTLE	GLDKLAIIDNIRLSDCSMEDAVOILQOCEDLVKLTKRDENSDSDRSGGALITYV	622
QY	592	PLND	SGSAGLVSVYKNSKRNHADLGIYKSIINGASAKDRLRVNDOLIAVNGESLL	651
DB	623	ELKR	YGGP-LGITISGT--EPPD-PIIISLTGKGLAERTGAIHIGRIILAINSSSLK	677
QY	652	GKT	ODAMETLRSMSTEGNKRGMTOIIVARRISKCNELKSPGPGPELPIETALDRE	711
DB	678	GKPL	SEDIHLQWAGER-----VTLKKIKOTQDAQPASSP--KKLPIP-----	717
QY	712	RLISH	SLYSGIEGLDESP--SRNALSR-----MGESGKYQLSPITYNM	753

DB	718	----	SHSD	LGGEEDPSPFIQPGKLSDVPSVPSAVDSWDSGIDARCGSGCTTFQ	774
QY	754	POD	VITIEDRLVLPPLHS--DOSSSSHDDVGFVTADACTWAKAISDSADCSLSPD	811	
DB	775	TSGN	FTYWRSPKKRASLSPVKKPPRSQTYPDVGLSND--WDRSYASGFAGASDSAD	831	
QY	812	VDPL	AFQREGF-----	831	
DB	832	AE	-----QEENFWSQALDELETCGSGILRELEATIMSGTMSLNHEAPFARQLGQAS	886	
QY	829	SEK	RQFSDASQLDEVKTRKSKM--DLGIADETKLTNYD--OKASPSRDVGPISGLK	885	
DB	887	FQER	SN-----SRPHYSQTRNSNTLPSPVGRKSVTLRKMGEIETMSPTPVELHKVTL	941	
QY	886	KSS	LSLQTAVAEVTLN-----CDLPFRHPRPRIIRGCGNESFRAAIDKS	932	
DB	942	KDS	MEDEFGFSVADGULEKGYVKNIRPAGPDGLGKLPYDRILQ--VNVHVRTRDFDCC	998	
QY	933	YDK	PAVDDEGMEITLEDTESSRSGRRESVSTASD-----QPSHSLEROMNQ	982	
DB	999	LVP	PLIASGNKLDLVISRNDIASQKSIQPALPDSWMSBQNSAFPOQPSH-----GGNL	1052	
QY	983	EKG	DKT	989	
DB	1053	ETRE	PTN	1059	

Search completed: July 24, 2002, 14:30:46  
 Job time: 7488 sec

OM of: US-09-757-781-2 to: EST:\* out\_format : pfs

Date: Jul 24, 2002 3:15 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

#### Command line parameters:

```
-MODEL=firmer+pn.model -DEV=xlh  
-O/cqgn2.1/USPTO.spool/US09757781/runtat.22072002.154234.17621/app_query.fasta_1.1434  
-DE-EST -OPMT=fastap -SUPERX=est -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.500 -CGAPOP=4.500  
-QGAPEXT=0.050 -YGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
-DEGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cd1  
-LIST=45 -DOCLAL=200 -THRX=score-pct -THRX=MAX=100 -THRX=MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09757781_ECGN1_1.3748  
-NCPU=6 -ICPU=3 -LONGLOS -DEV_TIMEOUT=120 -WARN_TIMEOUT=30  
-NO_XLPHY -WAIT -THREADS=1
```

#### Search information block:

Query: US-09-757-781-2

Query length: 1356

Database: EST:

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 2253.480000

score_list:	Strd Orig	ZScore	EScore	Len	Documentation	Mus musculus	adult me
Sequence							
gb_hlc:AK008055	+ 1678.00	1875.91	3.0e-95	2762	AK008055	Mus musculus	adult me
gb_est1:AL529863	- 1600.00	1796.78	7.8e-91	1078	AL529863	LIT_NFL001.N	
gb_est1:AL529864	- 1496.00	1670.69	2.6e-84	1074	AL529864	LIT_NFL001.N	
gb_est1:AL529864	+ 1317.50	1480.79	3.1e-73	852	BG745359	602723832FI NIH_MGC_11	
gb_est1:BM462581	+ 1307.00	1466.92	1.8e-72	1063	BM462581	AGENCOURT_6426569 NIH	
gb_est1:BM462581	+ 1296.00	1457.55	6.1e-72	767	B1917860	603183990FI NIH_MGC_12	
gb_est1:BM466811	+ 1241.00	1391.85	2.8e-68	1150	BM466811	AGENCOURT_6431318 NIH	
gb_est1:BM466811	+ 1216.00	1367.97	6.0e-67	725	BE792557	601585341FI NIH_MGC_7	
gb_est1:BM466811	- 1195.00	1340.46	1.9e-65	1040	BM466811	602723832FI NIH_MGC_1	
gb_est1:BM466811	+ 1181.00	1328.71	9.2e-65	713	BM466811	603322835FI NIH_MGC_89	
gb_est1:BM466811	+ 1150.00	1294.04	7.8e-63	694	BM466811	602805735FI NIH_MGC_8	
gb_est1:BM466811	+ 1146.00	1288.06	1.7e-62	814	BM466811	602805735FI NIH_MGC_8	
gb_est1:AL529864	+ 1132.00	1270.64	1.6e-61	973	AL529864	AU079784 Sugano mouse	
gb_est1:AL529864	+ 1127.50	1265.04	2.5e-61	830	AL529864	AL529864 LIT_NFL001.NH	
gb_est1:BM466811	+ 1118.50	1255.74	1.1e-60	941	BM466811	603202858FI NIH_MGC_97	
gb_est1:BM466811	+ 1118.00	1255.14	1.1e-60	945	BM466811	602434429FI NIH_MGC_46	
gb_est1:AL529865	+ 1109.00	1244.80	4.3e-60	966	AL529865	LIT_NFL001.NH	
gb_est1:BM466811	+ 1081.50	1213.78	2.3e-58	971	BM466811	602372744FI NIH_MGC_93	
gb_est1:BM466811	+ 1073.50	1204.64	7.5e-58	985	BM466811	602350296FI NIH_MGC_90	
gb_est1:BM466811	+ 1071.50	1205.56	6.6e-58	699	BM466811	602350296FI NIH_MGC_90	
gb_est1:BM466811	+ 1061.50	1193.82	3.0e-57	736	BM466811	602350296FI NIH_MGC_90	
gb_est1:BM466811	+ 1060.00	1189.41	5.2e-57	987	BM466811	602350296FI NIH_MGC_90	
gb_est1:BM466811	+ 1058.00	1188.58	5.2e-57	760	BM466811	602350296FI NIH_MGC_90	
gb_est1:BM466811	+ 1030.00	1159.98	2.3e-55	617	BM466811	602496888FI NIH_MGC_75	
gb_est1:BM466811	+ 1025.50	1148.69	9.8e-55	1207	BM466811	602093360FI NIH_MGC_16	
gb_est1:BM466811	+ 1025.00	1153.72	5.1e-55	660	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	+ 1016.50	1144.36	1.7e-54	645	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	- 1014.00	1144.10	2.3e-54	608	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	- 1013.00	1138.93	3.4e-54	758	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	+ 1001.00	1127.99	1.4e-53	574	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	+ 999.00	1125.45	9.9e-53	592	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	- 988.00	1112.95	1.9e-53	599	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	+ 976.50	1097.56	6.9e-52	779	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	+ 974.50	1096.12	8.3e-52	715	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	+ 964.00	1086.30	2.9e-51	754	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	- 963.50	1084.81	3.8e-51	678	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	- 943.50	1058.85	9.9e-50	921	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	+ 903.00	1017.65	1.9e-47	572	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	- 899.50	1011.33	4.4e-47	740	BM466811	602512525FI NIH_MGC_16	
gb_est1:BM466811	+ 898.50	1010.38	4.9e-47	725	BM466811	602512525FI NIH_MGC_16	

gb_est1:BM466811	+ 896.00	1007.65	7.0e-47	718	BM466811	RST26177	Altersys R
gb_est1:BM466811	- 893.00	1007.00	7.6e-47	535	BM466811	IL3-ET0116-281000-3	
gb_est1:BM466811	+ 893.00	1003.17	1.2e-46	809	BM466811	602860606FI NIH_MGC	
gb_est1:BM466811	- 882.50	993.02	4.6e-46	675	BM466811	B1083071	NIH_MGC
gb_est1:BM466811	+ 879.00	991.25	5.8e-46	534	BM466811	QVO-LR0014-250200-1	

seq\_name: gb\_hlc:AK008055

#### seq\_documentation\_block:

```
LOCUS AK008055 2762 bp mRNA linear HTC 19-JAN-2002  
DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length  
enriched library, clone:201002N16:homolog to PAR3, full insert  
sequence.  
ACCESSION AK008055  
VERSION AK008055.1 GI:12842001  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (strain: C57BL/6J) adult male small intestine cDNA to  
mRNA, clone:201002N16.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (sites)  
Carninci, P., and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2 (sites)  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3 (sites)  
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishi, K., Katsunaka, T., Tashiro, H., Itoh, M.,  
Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, K., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multiplexed sequencing  
Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
REFERENCE 4 (sites)  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
MEDLINE 11076861  
PUBMED 11076861
```

TITLE	JOURNAL	REFERENCE	AUTHORS
Functional annotation of a full-length mouse cDNA collection	Nature 409, 685-690 (2001)	5 (bases 1 to 2762)	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Butt, C., Carninci, P., Fukuda, S., Furukoshi, Y., Furuno, M., Hananaka, A., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirokawa, T., Horii, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasai, K., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koyama, S., Kutahara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numata, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamashita, T., Yasunishi, A., Yoshida, K., Yoshida, M., Muramatsu, M., and Hayashizaki, Y.
TITLE	JOURNAL	REFERENCE	AUTHORS
Submitted (10-JUN-2000) Yoshinori Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC).			



1183 TCGAGAACAAAGCAGACACCTCCTGCCCTGGGAACACAGCTCTGAAGCC 1232  
 416 pSerHisSerArgLeuProHisSerAlaHisProSerGlyLysProPro 433  
 1233 AGTACAT.....CTACACAGAGGAGCAGCCCTGAAGAGAGAGCCAG 1276  
 433 erAlaPro..... 435  
 1277 CCTCCCGCAGCAAGCAGAGTCCCGGGTCCAGACTAGAGAGAAA 1326  
 436 AlSerAlaProGlnAsnValPheSerThrValSerSerGlyTyrAs 452  
 1327 CCTTCCTCTCC.....TCACCTCTCTCTCTCATGGGGTTTGG 1364  
 452 nThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysGlyThrG 469  
 1365 CACCAAGAAAAATGCANAAAAATTAAGATGACCTAAAGAAAGGCCCTG 1414  
 469 IuclLysLeuGlyPheSerIleThrSerArgAspValThrIleGlyLys 485  
 1415 AGGACTTGGTTCTACTGTGTACCCAGAGATCTCTATACATGGTCT 1464  
 486 AlaProIleTyrValLysAsnIleLeuProArgLysAlaAlaIleGln 502  
 1465 GGTCCCATTTTGTAAAAACATCTTACCAAAAGAGCAGCAGTAAAGCA 1514  
 502 P..GlyArgLeuLysAlaGlyAspArgLeuIleGlnValAsnGlyValAsp 518  
 1515 TGGCC...CTACAAATCAGGGGACAGAAATTTGGAGTAAATGGCAGAGAT 1561  
 519 LeuValGlyLysSerGlnGlnGlnValValSerLeuLeuArgSerThrLys 535  
 1562 GTTACAGGAAGAACCAAGAAAGCTGTGGCATGTTAAGAGACACAA 1611  
 535 smetGlnGlyThrValSerLeuValPheArgGlnGlnAspAlaPheH 552  
 1612 GCAGGGAGACACAGCTGCATTTGCTCATTCGCCCGCAGAAAGGAGCTTCC 1661  
 552 IsProArgLysLeuAsnAlaGlnProSerGlnMetGlnIleProLysGln 568  
 1662 TGGCCGAGAGCTGAAGAGAACTGATTGTATGCTCTCCCTCGAG 1711  
 569 ThrLysAlaGlnAspGlnAspIleValLeuThrProAspGlyThrArgL 585  
 1712 TCAAGC.....GA 1719  
 585 uPheLeuThrPheGlnValProLeuAsnAspSerGlySerAlaGlyLeu 602  
 1720 GCAGCTCACCCTTGGAGATCCCTCGAATGATTTCTGGCTCCGCTGGCTGG 1769  
 602 LysValSerValLysGlyAsnArgSerLysGlnAsnHisAlaAspLeuGly 618  
 1770 GGGTTAGCCTGAAGAGGAATAGTCTCGAAMAACCTGCACTGACTTGGGG 1819  
 619 IlePheValLysSerIleIleAsnGlyLysAlaAlaSerLysAspGlyArg 635  
 1820 ATTTTATCAATCAATCATCATCTGAGGTGCTCTTCAAGATGAGCCG 1869  
 635 GlnArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerLeuG 652  
 1870 TCTCCCAATATATGACCACTGATGCTGTTAACGGGAAACTCTTCTGG 1919  
 652 LysThrAsnGlnAspAlaMetGlnThrLeuArgArgSerMetSerThr 668  
 1920 GAAAGTCAACCAACCAAGCATGGAACACTCAGGGCATTCATCTCATG 1969  
 669 GlnGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg.... 683  
 1970 GAGGAAACATCCGGGATGATCCAGTTGTATCTGAGAGAACCTGA 2019  
 684 .....IleSerLysCysAsnGlnLeuLysSerProGly 695  
 2020 GAGACCACTGGAGAGCTTTCAGAAATGTGAGACATTTCAGACCAAGCT 2069

695 erProProGlyProGlnLeuProIleGlnThrAlaLeuAspArgGln 711  
 2070 TTGAGAACTGTCAAGAA.....GCTGTGAGACCTCCAG 2104  
 712 ArgArgIleSerHisSerLeuTyrSerGlyIleGlnLysLeuAspGln 728  
 2105 CGAAATGATAGTACTATATATATAT..... 2128  
 728 rProSerArgAsnAlaAlaLeuSerArgIleMetGlyLysSerGlyLys 745  
 2129 .....CCGTTTGGCAGCT 2141  
 745 YrgLysLeuSerProThrValAsnMetProGlnAsp.....AspThr 758  
 2142 ATACT.....CCACAAAGTAAAGAGAAAGACCTA 2170  
 759 ValIle.....IleGlnAspAspArgLeuProValLeuPr 770  
 2171 TTGGTTCACAGTATGGGTGGCTGAGAAATGAACTACACCGTCCGCC 2220  
 770 oProHis.....LeuSerAspGlnSerSerSers 781  
 2221 ACCACATCTCTCTCGAATGGGCTTTGAAGATTTTCAGTCACAGCTCTG 2270  
 781 erHisAspAspValGlyPheValThrAlaAspAlaGlyThrTrpAlaLys 797  
 2271 GAGTGATACACACAGATATTT..... 2293  
 798 AlaAlaIleSerAspSerAlaAspCysSerLeuSerProAspValAspPr 814  
 2294 .....CCAGT..... 2299  
 814 oValLeuAlaPheGlnArgGlnGlyPheGlyArgGlnSerMetSerGln 831  
 2300 .....CAACATGCAACTTC..... 2314  
 831 ysArgThrLysGlnPheSerAspAlaSerGlnLeuAspPheValLysThr 847  
 2315 .....AGAACGTGACACACAGCTCAGGCGCCGAATGATTAACTT 2356  
 848 ArgLysSerLysSerMetAspLeuGlyIleAlaAspGlnThrLysLeuAs 864  
 2357 AAAGCTTCCAAGACATGAGACTT...GTGCCAATGCAAGCAAAAGTCA 2403  
 864 nThrValAspAspGlnLysAlaGlySerProSerArgAspValGlyPro 881  
 2404 GTCCGTGGCGGATGCGAGATCAGACTCTCCAGGCCAAGATTGTGTCAA 2453  
 881 erLeuGlyLeuLysLysSerSerSerLeuGlnSerLeuGlnThrAlaVal 897  
 2454 CTTTGGGTCTGAAGAACGCCAGTTCCTTGGAGACCTACAGACGGCTGTG 2503  
 898 AlaAlaValThrLeuAsnGlyAspIleProPheHisArgProArgProArg 914  
 2504 GCTGAACTCAGAGAAAT...GATCTGCCCTTCCACAGGCCCAACACACA 2550  
 914 GileIleArgGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspL 931  
 2551 CATGTGCGGGGCGGAGGCTCAACGAGAGCTTCCGAGCAGCCATCCACA 2600  
 2601 AATGCTACGATGGGCCAGAA...GAGGACAGCGGATGTGCTG..... 2641  
 948 LeuGlnGlnAspThrGlnGlnGlnSerSerArgSerLysArgGlnSerValSe 964  
 2642 .....TCTGATTAACAGCTCTCGCAGGCCACACAGCTTTG.. 2677  
 964 rThrAlaSerAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlyA 981  
 2678 .....AATTGTGAGTCTGCCCTCAGGGA 2702

```

    seq_name= gb_est1.AL529631
seq_documentation_block:
LOCUS       AL529631               1078 bp      mRNA           linear     EST_13-FEB-2001
DEFINITION   AL529631 LTR_NFL001_NB.C4 Homo sapiens cDNA clone CS0DD005YD18.3
              prime, mRNA sequence.
ACCESSION    AL529631
VERSION      AL529631.1 GI:12793124
SOURCE       EST.
            human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
             Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE    1 (bases 1 to 1078)
AUTHORS      Li W.B., Gruber,C., Jesssee,J. and Polayes,D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Source
Location/Qualifiers
1..1078
 /organism="Homo sapiens"
/db_xref="taxon:9606"
/cloned="CSODD005YD18"
/clone_id="LTR_NFL001_NB.C4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA was primed with a NotI-o.ligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT  236 a          222 g         331 t                6 others
ORIGIN
alignment_scores:
Quality: 1600.00 Length: 359 Gaps: 4
Percent Similarity: 93.315 Percent Identity: 90.251
Alignment_block:
US-09-757-781-2 x AL529631/review ..
Align seg 1/1 to reverse of: AL529631 from: 1 to: 1078
696 ProProgliprogileuproliegltThralaleuspaspargluar 712
|||||
1063 CCCCCCTGGACTGCGTCATGAACAAGCCGTGTGATATAGAANAAGC 1014
|||||
712 GARGLseerHisserLeutyrseryglylegluglyleuaspgluserp 729
|||||
1013 AAGAAATTCCCATTTCCCTCAAGTGAGATTGAGGGCTTGATGATGCCGC 964
|||||
729 roSerArgasnalalaAleuseArgileweetlgylusergluystryr 745
|||||
963 CCAGCAAATAATGCGCCCTCAGTAGAATAATC.....GATAAATTC 923
```

---

-----

```

746 GLInleuserProthrValAsnMetProglinspsphrrvalilleelg 762
|||||
922 CRGCTGTCCCCTCACAGTAATATATGCCCCAACATGACACTGTCAATAGA 873
|||||
762 uASpAspragleuproVallieuproothislesuserAspginusers 779
|||||
872 AGATGACAGGTTGGCCAGTCTTCCACCACATCTCTGACAGCTCCCTT 823
|||||
779 erSeSerHiAsPASPvAlgiPhyheValThrAlaaPAalaglyhtTP 795
|||||
822 CCSGTCOCATGATGATGATGGGGTTTTGTGACGGAGATGCTGTGACTTGG 773
|||||
796 AlalySaialaalieserSperrlaaSPcyserleuserProspya 812
|||||
772 GCCAAGGCTGCATACATGATTCAGCCGACTGCTTTAGTGCAGATGT 723
|||||
812 lASproValleualaphegninragugliglyphegyilaraginsemets 829
|||||
722 TGATCCAGTCTTCTTCTTTTAAcAgagaGaatttggAcctCaGAGATGT 673
|||||
829 ergLulsArgrthyrgLnPheserAsPaalerGlneuaspHeVAL 845
672 CAGAAAAACGCAAGAACAATTTTCAGATGCCAGTCAATTCGATTCGT 623
|||||
846 LystrhyrglyserryssermetaspleugilyllealaspluthrLy 862
|||||
622 AAAACAGAAAATCAAAAAGCATGATTTA...gttgacagacaGTAA 576
|||||
862 sLeuaSrthraLaspaSplnllyaLaigyserProseratrgaspVLg 879
575 ACYCAATATACAGTGATGACAGAGAAGCAAGTCTCCMACAGAGTGG 526
|||||
879 lyProserleunglyleulylsyserserSerleuglisertleugnathr 895
525 GTCTTCTCCGTGGGTGAAGAAGTCAAGCTGTTGAGAGCTGCACAGCC 476
|||||
896 AlaVaLAJaLuvaThrlenusngilyspieleropheniSAAgrPror 912
|||||
475 GCAGTTGCCGAGGTGACTTGAATGGGATATTCCTTCCAACYCTCACG 426
|||||
912 ghroArghllleahrglylarhglyCyasaNguluserPheargraALai 929
425 GGCGGGATAAATCAAGCAGGGGATGCAAAAGRCCTTCAGGCTGCCA 376
|||||
929 laSPlyseSTyrzAspySProlavaLaspaSPAappsgulgymet 945
375 TCAGCAAAATCTTATGATTAACCCCGGTAGATGATGATGAAGGCATG 326
|||||
946 GlutrhreuglugluaspThrnglgluseserargserylarginuse 962
|||||
325 GAGACTTGGAAAGAACACAGAAAGAAAGTTCAAGATCAGGAGAGAGTC 276
|||||
275 TGTATCCACAGCCAGTATCAGCTTCCACTCTCTGAGAGACAATAGA 226
|||||
979 snGliYaenGingluNygLYAsplyStrAsparqlysyAsapltyTr 995
|||||
225 ATGGAAACCAAGAGAAAGGTGATPAAGCTGATGAAGAAAAAGATAAACT 176
|||||
996 GlySySLuLySuLySaPaPygaPYSGULyuLyAsPLyskelysaL 1012
|||||
175 GGAHAHAHAHAHAHAHGATNAKNTAAAGGAAGAAGATAAATGAAGC 126
|||||
1012 alYSlySYGLYMETLEUlySGLyeUGlylaspetMearghpehiLYSH 1029
125 CAAGAAAGCAATCTCAAGAGGCTTGGAGACATAGTTCAGCCCTGCCAA 77
|||||
1029 IsArqlySaPasPlYSIllegLUlysrThChlyLSyllleystInglnu 1045
||| .....|||
76 . CTGAAGCCCGAGAAAGACATGACAACAGCAAGCAATTCAAA..... 38
```



1046 SerptherSergluGluluarGlie 1054  
|||||  
37 .....ACATGTTTGACAGCACATA 17

seq\_name: gb\_est1:AL529864

seq\_documentation\_block:  
LOCUS AL529864 1074 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL529864 LTI\_NFL001\_NBC4 Homo sapiens cDNA clone CSDD005YH02 3  
prime, mRNA sequence.  
ACCESSION AL529864  
VERSION AL529864.1 GI:12793357  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1074)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
source  
1..1074  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CSDD005YH02"  
/clone\_lib="LTI\_NFL001\_NBC4"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
vector. Library was normalized. Library was constructed  
by Life Technologies. Contact : Feng Liang Life  
Technologies, a division of Invitrogen 9800 Medical Center  
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
8371 Email : liliang@life.com  
http://fulllength.invitrogen.com"

BASE COUNT 229 a 289 c 221 g 324 t 11 others  
ORIGIN

alignment\_scores:  
Quality: 1496.00 Length: 347  
Ratio: 4.603 Gaps: 4  
Percent Similarity: 93.660 Percent Identity: 88.473

alignment\_block:  
US-09-757-781-2 x AL529864/rev ..

Align seg 1/1 to reverse of: AL529864 from: 1 to: 1074

696 ProProGlyProGluLeuProIleGluThrAlaLeuAspAspArgGluAr 712  
||||| ||| ||||| :|||||  
1042 CCCCCCTGACCTGGTCCCATTTG...ACACGCTGATGATGAGACAGC 997  
712 gatgileserHisSerLeuTySerGlyIleGluGlyLeuAspGluSerP 729  
||||| :|||||  
996 .ACAAATTCCTTCTTACAGTGGAGTTGAGGGCTTGATGAATCGC 947  
729 roserarGanaAlaAlaLeuSerArgIleMetGlyGluSerGlyLysTyr 745  
||||| :|||||  
946 CCAGCAGAAATGCTGCCCTCAGTAGATTAATG.....GGTAATAC 906  
746 GluLeuSerProThrValaAsnMetProGluAspAspThrValaIleIleG 762  
||||| :|||||  
905 CMGCTGCTCCCTACAGTAATATGCCCCCAAGATGACACTGTCTATTAGA 856

762 uaAspAspArgLeuProValaLeuProProHisLeuSerAspGlnSerSers 779  
||||| :|||||  
855 AGATGACAGAGTTCACAGTCTCTCCACATCTCTGACACAGCTCTTCT 806  
779 erSerSerHisAspAspValGlyPheValThrAlaAspAlaGlyThrTrp 795  
||||| :|||||  
805 CCVAGCTCCCTGATGATGTTGGGTTTGTGACGAGCATGCTGTGACTTGG 756  
796 AlAlaValAlaAlaIleSerAspSerAlaAspCysSerLeuSerProAspVa 812  
||||| :|||||  
755 GCCAAGGCTCAATCACTGATGATGACCGGACTGCTTTGATGATCCAGATGT 706  
812 laAspProValaLeuAlaPheGlnArgGluGlyPheGlyArgGlnSerMetS 829  
||||| :|||||  
705 TGATCCAGTCTCTGCTTTTCACAGAGAATTTGACGCTCAGAGTGTGT 656  
829 erGluLysArgThrLysGlnPheSerAspAlaSerGlnLeuAspPheVal 845  
||||| :|||||  
655 CAGAAACGACACAAAGCAATTTTCAGATGCCACTCAATGGATTTGCTT 606  
846 LysThrArgLysSerLysSerMetAspLeuGlyIleAlaAspGluThrLys 862  
||||| :|||||  
605 AAAACACGAAATCAAAAACCATGAT...TTWCTTCTGACGAGACTTA 559  
558 ACTCAATACAGTGGATGACCAAGAAAGCAGGCTTCCCGCAGACATGTGG 509  
879 LysProSerLeuGlyLeuLysSerSerSerLeuGlnSerLeuGlnThr 895  
||||| :|||||  
508 GTCTCTCTCCCTGGGCTGTAAGAAAGCTGCTGTTGGAGAGCTGCAACAC 459  
896 AlAlaValaGluValaThrLeuAsnGlyAspIleProPheHisArgProAr 912  
||||| :|||||  
458 GCAGTTCCTCCAGAGTGAATGAGGATATCTCTTCCATCTCCACAGC 409  
912 gProArgIleIleArgGlyArgGlyCysAsnGlnSerPheArgAlaAlaI 929  
||||| :|||||  
408 GCCCGGATATATCAGAGCAGAGGATGCAATGACGCTTCAGGCTGCCA 359  
929 leAspLysSerTyrAspLysProAlaValaAspAspAspGluGlyMet 945  
||||| :|||||  
358 TCGCAATCTTGTGATTAACCCCGGCTAGATGATGATGAAGAGCATG 309  
946 GluThrLeuGluGluAspThrGluGluSerSerArgSerGlyArgGluSe 962  
||||| :|||||  
308 GAGACCTTGGAGAGACACAGAAAGATTCAGATTCAGGAGAGATC 259  
962 rValSerThrAlaSerAspGlnProSerHisSerLeuGluArgGlnMetA 979  
||||| :|||||  
258 TGTATCCACAGCCAGTATCAGCTTCCACTCTCTGAGAGCAATAATGA 209  
979 snGlyAsnGlnGluLysGlyAspLysThrAspArgLysLysAspLysThr 995  
||||| :|||||  
208 ATGGAAACCAAGAGAAAGGTGATTAAGACTGATTAACAAAAGATTAAGACT 159  
996 GlyLysGluLysLysLysAspArgAspLysGluLysAspLysMetLysAl 1012  
||||| :|||||  
158 GGAAAAAGAAAAAGAAAGATAGAGATAGAGAGAAAGATTAATAAGAAC 109  
1012 alLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArgPheGlyLysH 1029  
||||| :|||||  
108 CAAGAAAGGAATGCTGAAGGCTTGGGAGACATGTTCCACCTTGCCAAA. 60  
1029 lsArgLysAspAspLysIleGluLysThrGlyLysIleLys 1042  
||||| :|||||  
59 ..CTGAAGCCCGAGAGAGATGACAAACAAAGCATTCAAA 21

seq\_name: gb\_est2:BG745599

seq\_documentation\_block:  
LOCUS BG745599 852 bp mRNA linear EST 15-MAY-2001

```

DEFINITION      602723832f1 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:4850526 5',
                  mRNA sequence.
ACCESSION       BG745599
VERSION         BG745599.1 GI:14056252
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS         National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE           Unpublished (1999)
JOURNAL         Contact: Robert Strausberg, Ph.D.
COMMENT         Email: cgabs-remail.nih.gov
                  Tissue Procurement: Dr. Mark Watson
                  CDNA Library Preparation: Ling Hong/Rubin Laboratory
                  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
                  DNA Sequencing by: Incyte Genomics, Inc.
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LNL at:
                  http://image.lnl.gov
                  Plate: LICM691 row: 0 column: 07
                  High quality sequence stop: 829.
FEATURES
  source         1..852
                  Location/Qualifiers
                  1..852
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4850526"
                  /clone_lib="NIH_MGC_113"
                  /lab_host="DH10B (phage-resistant)"
                  /note="Organ: Spleen; Vector: pOTB7; Site:1: XhoI; Site:2:
                  EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                  into EcoRI/XhoI sites using the following 5' adaptor:
                  GGCAAGAG(G). Library constructed by Ling Hong in the
                  Laboratory of Gerald M. Rubin (University of California,
                  Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                  Superscript II RT (Life Technologies). Note: this is a
                  NIH-MGC Library."
BASE COUNT      237 a 191 c 222 g 202 t
ORIGIN
alignment_scores:
  Quality: 1317.50      Length: 300
  Ratio: 4.739          Gaps: 6
  Percent Similarity: 92.667 Percent Identity: 90.667
alignment_block:
  US-09-757-781-2 x BG745599 ..
Align seg 1/1 to: BG745599 from: 1 to: 852
649 SerLeuLeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSe 665
    ::::::::::::::::::::::::::::::::::::::::::::::::::::
    1 GCCCGTGGGGGCAAGCAAAACCAAGATGCGCGGAAACCCCTTAAGAGGTC 50
665 rMetSerThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAla 682
    |||||||
    51 TATGTCCTACTGAAGCATTAACGAGAAATGATCCAGCTTATGTTGCCA 100
682 rGArgIleSerLysCysAsnGlnLeuLysSerProGlySerProGly 698
    |||||||
    101 GGAGAAATTAAGCAATGCAATGAGTCACTGGAGCCCTTCGGA 150
699 ProGluLeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSe 715
    |||||||
    151 CCTGAGCTGCCCATTTGAACACAGCTTGATGTACAGAAAGAAATTC 200
715 rHisSerLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArg 732
    |||||||
    201 CCATTCCTCTACACTGGGATTTGAGGGGCTTGATCAATGCCACCACAGA 250

```

```

732 snAlaAlaLeuSerArgIleMetGlyIleSerGlyLysTyrGlnLeuSer 748
    |||||||
    251 ATGTCGCCCTCAGTAGATTAATG.....GGTAAATACACAGCTGTCC 291
749 ProThrValAsnMetProGlnAspAspThrValIleIleGluAspAsp 765
    |||||||
    292 CCTACAGTGAATATGCCCCAGATGACACTGTCAATTAAGAAGAGACAG 341
765 GluProValLeuProProHisLeuSerAspGlnSerSerSerSer 782
    |||||||
    342 GTTGCAGTGTCTTCTCCACATCTCTGACGACGCTCTTCCAGCTGCC 391
782 IsAspAspValGlyPheValThrAlaAspAlaGlyThrTyrAlaLysAla 798
    |||||||
    392 ATGATGATGTGGGGTTGTGACGCGAGATGCTGTACTGGGCCAAGGCT 441
799 AlaIleSerAspSerAlaAspCysSerLeuSerProAspValAspPro 815
    |||||||
    442 GCAATCACTGATTCAGCCGACACTGCTTTGAGTCCAGATGTTGATCCAG 491
815 IleuAlaPheGlnArgGluGlyPheGlyArgGlnSerMetSerGlyLys 832
    |||||||
    492 TCTTCTCTTTACACGAGAAAGATTTGGACGTCAAGTATGTCAGAAAAAC 541
832 rGThrLysGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArg 848
    |||||||
    542 GCACAAGCAATTTTCAGATGCCAGTCAATTCGATTTGTTAAACACAGA 591
849 LysSerLysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsn 865
    |||||||
    592 AAATTAAGCAATGATGATTTAGT..... 615
865 rValAspAspGlnLysAlaGlySerProSerArgAspValGlyProSer 882
    ::::::::::::::::::::::AGTTCTCCACAGAGATGGGCTCTTCCC 646
616 ..... 646
882 euGlyLeuLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAla 898
    |||||||
    647 TGGGCTCTAAGAAAGTCAAGCTCATTTGAGAGTCTCAACACCCAGTTGCC 696
899 GluValThrLeuAsnGlyAspIlePro.PheHisArgProArgProArg 914
    |||||||
    697 CGAGGTGACTTTGAATGGGAGATATTCCTTTTCCATGCTCCAGCGCGG 746
915 IleIleArgGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAsp 931
    |||||||
    747 ATCTTCAGAGCGCAGGAGATGCAATGAGAGCTTCAGAGCTGCCATCGACA 796
931 sSerTyrAspLysPro.AlalaAspAspAspAspGluGlyMet 945
    |||||||
    797 ATCTTATGATTAACCGCGGT...AATGATGATGATGAAGGACATG 837
seq_name: gb_est2:BM462581
seq_documentation_block:
  LOCUS      BM462581 1063 bp mRNA linear EST 05-FEB-2002
  DEFINITION AGENCOURT_6426369 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5518290
  5', mRNA sequence.
  ACCESSION  BM462581
  VERSION    BM462581.1 GI:18511621
  KEYWORDS   EST.
  SOURCE     human.
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS      National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE        Unpublished (1999)
JOURNAL      Contact: Robert Strausberg, Ph.D.
COMMENT      Email: cgabs-remail.nih.gov
              Tissue Procurement: ATCC
              CDNA Library Preparation: Life Technologies, Inc.

```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LLAM12178 row: n column: 19  
 High quality sequence stop: 365.

## FEATURES

source  
 Location/Qualifiers  
 1. 1063  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="5318290"  
 /clone\_lib="NIH\_MGC\_71"  
 /tissue\_type="Telomysarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: uterus; Vector: PCMV-SPORE; Site: 1; Note:  
 Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 2.1 kb."  
 BASE COUNT 307 a 247 c 259 g 250 t  
 ORIGIN

alignment\_scores:  
 Quality: 1307.00 Length: 363  
 Ratio: 4.059 Gaps: 16  
 Percent Similarity: 88.705 Percent Identity: 84.298

alignment\_block:  
 US-09-757-781-2 x BM462581 ..

Align seg 1/1 to: BM462581 from: 1 to: 1063

```

575 AspIleValIleuThrProAspGlyThrArgGluPheLeuThrPheGluVal 591
|||||
1 GATATTGTTCTTACACCTGATGCGACAGGAGATTCTGCATTGAGT 50
591 lProLeuAsnAspSerGlySerAlaGlyLeuGlyValSerValLysGlyA 608
|||||
51 CCCACTTAATGATTCAGATCTGACAGCCCTGGTGTCTCAGTCAAAAGCTA 100
608 snArgSerLysGluAsnHisAlaAspLeuGlyLLePheValLysSerIle 624
|||||
101 ACCGGCTAAAGGAAACACGCGATTTGGGAATCTTTGTCAAGTCCATT 150
625 IleAsnGlyValAlaAserLysAspGlyArgLeuArgValAsnAspGly 641
|||||
151 ATTAATGAGAGACGACATCTAAAGATGGAAGGCTCGGGTGAATGATCA 200
641 nIleuIleAlaValAsnGlyLysSerLeuLeuGlyLysThrAsnGlnAspA 658
|||||
201 ACTGATAGCAGTAATGAGAAATCCCTGTGGCGCAACCAACCAATATG 250
658 lAmetsLurIleuArgArgSerMetSerThrGluGlyAsnLysArgGly 674
|||||
251 CCATGGAAACCTTAAGAGGTCTATGTCTACTAGGCAATTAACGAGGA 300
675 MetIleGlnLeuIleValAlaArgArgIleSerLysCysAsnGluLeuGly 691
|||||
301 ATGATCCAGCTTATGTGCAAGAGAAATTAAGCAAGTCAATGAGCGAA 350
691 sSerProGlySerProProGlyProGluLeuProIleGluThrAlaLeuA 708
|||||
351 GTCCACCTGGAGCCCTCGACCTGACCTGAGCTGCAATTAAGACGCTTG 400
708 spAspArgGluArgArgIleSerHisSerLeuThrSerGlyIleGluGly 724
|||||
401 ATGATGAGAGACGAAATTTCCCATTTCCCTCTACAGTGGGATGAGGG 450
725 lLeuAspGlySerProSerArgAsnAlaAlaLeuSerArgIleMetGlyGly 741
|||||
451 CTGATGAATACGCCAGACAGAAATGCGCCCTCAGTAGATTAATG.... 495
741 uSerGlyLysTyrGlnLeuSerProThrValAsnMetProGlnAspAsp 758

```

```

496 .....GGTAAATACCAAGCTGTCCTTACAGTGAATATCCCAAGATGACA 541
758 hrValIleIleGluAspArgLeuProValIleuProProHisLeuSer 774
|||||
542 CTGCTATTATAGAGATGACAGGTGGCAGTGTCTCTCCACATCTCTCT 591
775 AspGlnSerSerSerSerHisAspAspValGlyPheValThrAlaAs 791
592 GACACAGTCCCTTCCACCTCCCATGATGATGTGGGTTTGTAGCGGAGA 641
791 palAGlyThrTrrPalAlaLysAlaIleSerAspSerAlaAspCysSerL 808
642 TGGTGTGACTTGGGCCCAAGCT.GCAATCAGTGAATTCAGCCGACTGCTT 690
808 euserProAspValAspProValLeuAla.PheGlnArgGluGlyPheG 824
691 TTGAGTCCGATGGTGAATCAGTCTTGTCTTTCACAGAAAGATTTGG 740
824 YArgGlnSerMetSerGlu.LysArgThrLysGln.PheSerAspAla.S 840
741 ACGTGACAGTATGTCAGAAAAAAGCACAAGCAATTTCCAGATGCCCA 790
840 erGlnLeuAspPheVal.LysThrArgLysSerLysSerMetLasp.LeuG 856
791 GCCAATTTGATTTCTTTAAACCCCAAAATCCAAAAGCCTGGAAATTAG 840
856 lylIleAla.AspGluThrLysLeuAsnThrValAsp..AspGlnLysAla 871
841 GGATAGCTTACGAGACTAAATCTCATATCCTGGATGCGCCCAAAACCA 890
872 GlySerProSerArgAspVal...GlyProSerLeuGlyLeuLysLys 887
891 GGCTCTCCCAACCAAAATGTGGGCGCTTCTTGGGTCTGCAAAAAA 940
887 rSerSerLeu.....GluSerLeu.GlnThr.....AlaValAla 898
941 GTCAMACCTCTTGAAGAAAGTGTGGCAAAACCCCAATTTGGCCCAAGG 990
899 GluValThrLeuAsnGlyAspIleProPheHis...ArgProArg...Pr 913
991 GCACCTTTAAAGGGGAAAAATTCCTTTTCAATGCGGCCCAAGGCC 1040
913 oArgIleIle...ArgGlyArg 919
1041 GGGGATTAATTCACCAAGGCGCG 1062
seq_name: gb_est2:BI917860
seq_documentation_block:
LOCUS BI917860 767 bp mRNA linear EST 16-OCT-2001
DEFINITION 60318390F1 NIH_MGC_121 Homo sapiens CDNA clone IMAGE:5247718 5',
mRNA sequence.
ACCESSION BI917860
VERSION BI917860.1 GI:16199788
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 767)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

Plate: ILAM11625 row: d column: 23  
 High quality sequence stop: 759.  
 Location/Qualifiers

## FEATURES

source

1. 767  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5247718"  
 /clone\_11b="NIH\_MGC\_121"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH\_MGC library."  
 BASE COUNT 236 a 179 c 190 g 162 t  
 ORIGIN

## alignment\_scores:

Quality: 1296.00 Length: 255  
 Ratio: 5.082 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.608

## alignment\_block:

US-09-757-781-2 x BI917860 ..

Align seg 1/1 to: BI917860 from: 1 to: 767

373 GULYASASANTYRYSERARPHESERPROASPERGLNTRYL 389  
 |||||||||||||||||||||||||||||||||||||||||||  
 3 GAGAGAAACAATTACTATTCAAGCCGTTTACCCGACAGCCAGTAT 52  
 |||||||||||||||||||||||||||||||||||||||||||  
 389 GASPASATSERVALASNSERIALGLYUHSRTHVALGINATGALAP 406  
 |||||||||||||||||||||||||||||||||||||||||||  
 53 TCACAAACAGAGTGTGAACAGTGCAGGGCTTCACACGGTGACAGAGC 102  
 |||||||||||||||||||||||||||||||||||||||||||  
 406 TQATGLEUASNHISPROPROGLUGLNLIASPSERHISERARGLEUPRO 422  
 |||||||||||||||||||||||||||||||||||||||||||  
 103 CCCGACTGACACACCCGCTGAGCAGATGACTCTCACAAGACACTACCT 152  
 |||||||||||||||||||||||||||||||||||||||||||  
 423 HISSEALAHISPROSERGLYSPROPSERALAPROLASERLAPR 439  
 |||||||||||||||||||||||||||||||||||||||||||  
 153 CATAGGCGACACCCCTCGGGAAAAACCACTCCGCTCCAGCCTCGGACCC 202  
 |||||||||||||||||||||||||||||||||||||||||||  
 439 OGINSANVALPHESETRTHTRVALSERSEGLYTRASNTHRLYLSI 456  
 |||||||||||||||||||||||||||||||||||||||||||  
 203 TCAGATGATTTAGTAGACACTGTAAGCAGTGTTATACACCAAAAAA 252  
 |||||||||||||||||||||||||||||||||||||||||||  
 456 IAGGLYLSARGLEUASNLIEGLINLEULYSLGLYTHRGUUGLYLEUGLY 472  
 |||||||||||||||||||||||||||||||||||||||||||  
 253 TAGGCAAGAGCGCTTATATCCAGCTTAAGAAAGTACACAAAGTTTGGGA 302  
 |||||||||||||||||||||||||||||||||||||||||||  
 473 PHESETRTHTRSERARGSPVALTHTRILEGLYGLYSERLAPROTIETY 489  
 |||||||||||||||||||||||||||||||||||||||||||  
 303 TTCAGCATCACTCCAGAGATGTAAACAATAGGTGGTCCAGCTCCAAATCTA 352  
 |||||||||||||||||||||||||||||||||||||||||||  
 489 TVALYASNLIELEUPROARGLYALALALIEGLINASPGLYARGLEUL 506  
 |||||||||||||||||||||||||||||||||||||||||||  
 353 TGTGAAAAAATTCCTCCCGGGGGGGCGCATTCAGATGGCGGACTTA 402  
 |||||||||||||||||||||||||||||||||||||||||||  
 506 YSALAGLYSPARGLEULIEGLIVALASNLGLYVALASPLEUVALGLYLS 522  
 |||||||||||||||||||||||||||||||||||||||||||  
 403 AGGCAGAGACAGACTTATAGAGTAAATGAGTAGATTAGGGGCAAA 452  
 |||||||||||||||||||||||||||||||||||||||||||  
 523 SERGLNGUGLUVALSERLEULEUARGSERTHLYMERGLUGLYTH 539  
 |||||||||||||||||||||||||||||||||||||||||||  
 453 TCCCAAGAGAGAGATTGTTGGCTTGAGAGACCAACAAGATGAAGAAC 502

539 TVALSERLEULEUVALPHEARGINGLUASPALAHEHISPROARGLIUL 556  
 |||||||||||||||||||||||||||||||||||||||||||  
 503 TGTGAGCCCTTCGGCTTCGCCAGAGAGAGCCCTTCACCAAGAGAAC 552  
 |||||||||||||||||||||||||||||||||||||||||||  
 556 EULASALAGLUPROSERGLINMETGINLIEPROLYSGLUTHTHLYSLAGLU 572  
 |||||||||||||||||||||||||||||||||||||||||||  
 553 TGAATGACAGACCAAGCAGATGAGATTCAGAAAAAGAAACAGAAACAGAA 602  
 |||||||||||||||||||||||||||||||||||||||||||  
 573 ASPGLUASPILEVALLEUTHTRPROASPLGTHTRARGLUPHELEUTHPH 589  
 |||||||||||||||||||||||||||||||||||||||||||  
 603 GATGAGGATATCGCTTCTTACACCTGATGACACAGAGAAATTCGACAT 652  
 |||||||||||||||||||||||||||||||||||||||||||  
 589 EGLUVALPROLEUASNASPSERGLYSERIALGLYLEUGLYVALSERVAL 606  
 |||||||||||||||||||||||||||||||||||||||||||  
 653 TGAATCCCACTTATGATTCAGATGTCAGAGCCCTGTGTGCACTGTCA 702  
 |||||||||||||||||||||||||||||||||||||||||||  
 606 YSGLYASNARGSERLYSGLUASNHISALAPLEUGLYLIEPHEVALYS 622  
 |||||||||||||||||||||||||||||||||||||||||||  
 703 AAGGTAAACCGGTCAAAAGAACACGACGAGATCTGGGAATCTTGTCAG 752  
 |||||||||||||||||||||||||||||||||||||||||||  
 623 SERILEIASNLGLY 627  
 |||||||||||||||||||||||||||||||||||||||||||  
 753 TCCATTATTACTCGG 767

seq\_name: gb\_esc2:BM466811

seq\_documentation\_block:

LOCUS BM466811

DEFINITION AGENCOURT\_6431318 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:5503072

5', mRNA sequence.

ACCESSION BM466811

VERSION BM466811.1 GI:18515853

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE NIH-MGC http://mgs.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: egadps-femail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: ILAM12142 row: d column: 17

High quality sequence start: 9

High quality sequence stop: 652.

## FEATURES

source

Location/Qualifiers  
 1. 1150  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5503072"  
 /clone\_11b="NIH\_MGC\_67"  
 /tissue\_type="retinoblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."  
 BASE COUNT 294 a 310 c 309 g 235 t 2 others  
 ORIGIN

## alignment\_scores:

Quality: 1241.00 Length: 246  
 Ratio: 5.107 Gaps: 3  
 Percent Similarity: 98.780 Percent Identity: 98.374





440 CAGGGATCAATGAGAGCTTCAGAGCTGCCATCGCAAAATCTTATGATA 391  
935 yspRoAlaValAspAspAspAspGluGlyMetGluThrLeuGluGluAsp 951  
|||||  
390 AACCCCGGTAGATGATGATGATGAAGGATGAGACCTTGGAAGAAAGAC 341  
952 ThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlaSerAs 968  
|||||  
340 ACAGAGAAAGTTCAAGATCAGGAGAGAGATCTGTATCCACAGCAGTGA 291  
968 pcInProSerHisSerLeuGluArgGlnMetAsnGlyAsnGlnGluysG 965  
|||||  
290 TCAGCTTCCACTCTCTGGAGACAAATGAATGGAACCAAGAGAGAAAG 241  
985 lYAspLyThrAspArgLyLysAspLySerThrGlySerGluLysLyLys 1001  
|||||  
240 GTGATTAAGACTGATAGAAAAAGGATTAATACTGAAAAAGAAAGAAAG 191  
1002 AspArgAspLySerGluLysAspLySerMetLysAlaLysLySerGlyMetLeu 1018  
|||||  
190 GATAGAGATTAAGAGAGAGATTAATACTGAAAAAGCAAGAGGATGCTGAA 141  
1018 sGlyLeuGlyAspMetPheArgPheGlyLys 1028  
|||||  
140 GGGCTTGGAGACATGTCTAGCCTTGCCAAA 110  
seq\_name: gb\_est2: BG170486  
seq\_documentation\_block:  
LOCUS BG170486 713 bp mRNA linear EST 06-FEB-2001  
DEFINITION 60322859F1 NIH\_MGC\_89 Homo sapiens cDNA clone IMAGE:4425975 5',  
mRNA sequence.  
ACCESSION BG170486  
VERSION BG170486.1 GI:12677189  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 713)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-rt@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10172 row: e column: 16  
High quality sequence stop: 710.  
Location/Qualifiers  
1..713  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4425975"  
/clone\_lib="NIH\_MGC\_89"  
/tissue\_type="hypertrophoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pCMV-SPORT6; site: 1; NotI;  
site: 2; SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."  
BASE COUNT 213 a 167 c 206 g 127 t  
ORIGIN  
alignment\_scores: Quality: 1181.00 Length: 235

Ratio: 5.069 Gaps: 0  
Percent Similarity: 99.149 Percent Identity: 99.149  
alignment\_block:  
US-09-757-781-2 x BG170486 ..  
Align seg 1/1 to: BG170486 from: 1 to: 713  
1034 lYsAlleGluysThrGlyLysIleValIleGlnGluSerPheThrSerG 1050  
|||||  
2 AAGATTAGAAACGGGTAAATTAATTAATACGAAATCTTTACATCA 51  
1050 uGluGluArgIleArgMetLysGlnGluGluArgIleGlnAlaLysT 1067  
52 AGAGAGAGAGATAGCAATGAAGACAGAGACAGAGATTAACCCAAA 101  
1067 hrArgGluPheArgGluArgGlnAlaArgGluArgAspTyrAlaGluIle 1083  
|||||  
102 CTCGAGAAATTTAGGGAACGACACAGCTCGAGCGCTGATGCTGAAT 151  
1084 GluAspPheHisArgThrPheGlyCysAspAspGluMetLysGly 1100  
|||||  
152 CAAGATTTTCATCGGACATTTGGCTGTGATGATGATGATGATGATG 201  
1100 yValSerSerTyrGluGlySerMetAlaLeuAsnAlaArgProGlnSer 1117  
202 AGTTTCTTCTATGAGGTTCATGGCTTCATGGCTTCAGCTAGACTCAG 251  
1117 roArgGluGlyHisMetMetAspAlaLeuTyrAlaGlnValLysPro 1133  
252 CACAGAGAGGCGCATATGATGATGATGATGATGATGATGATGATG 300  
1134 ArgAsnSerLysProSerProValAspSerAsnArgSerThrProSerAs 1150  
301 CGGAATTCAAAACCTCACTCCTGATGACATGATGATGATGATGATG 350  
1150 nHisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGln 1167  
351 TCATGATTCGATGATACAGCGTCTGAGGCAAGATTTGACAGCAAGAG 400  
1167 spGluAspValGluAspArgArgArgThrTyrSerPheGlnGlnProT 1183  
401 ATGAAGATGTGAAGATGCTCGCGGACCTTACTTTGACCAACCTCG 450  
1184 ProAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerVal 1200  
451 CCGAAGCGACGCGCGCGGACGACAGCGGCGGACACTGCTCGGTGA 500  
1200 uValGlnMetGlnArgGlnArgGlnGluArgGluArgLysSerGlnGln 1217  
501 GTGTCAGATGAGCGGCGGCGGAGAGAGAGCGGAGAGCTCCAGAGAG 550  
1217 laGlnArgGlnTyrSerSerLeuProArgGlnSerArgLysAsnAlaSer 1233  
551 CCCAGCCCACTACAGCTCTGCTCGGCAACAGAGAAAAAGCCAGC 600  
1234 SerValSerGlnAspSerTyrGluGlnAlaProTyrSerProGlyGlu 1250  
601 TCGGTCTCCCAAGACTCTTGAGACAGAACTACTCCCTTGAGGAAGCT 650  
1250 eGlnSerAlaLysGluAsnProArgTyrSerSerTyrGlnGlnLysArg 1267  
651 CCAAGATGCGCAAGAGAAAGAACCCAGAGTACTCCAGGCTACAGAGCT 700  
1267 snGly 1268  
|||||  
701 ACGGT 705  
seq\_name: gb\_est2: BG910244  
seq\_documentation\_block:  
LOCUS BG910244 694 bp mRNA linear EST 05-JUN-2001  
DEFINITION 602805735F1 NCI\_CGAP\_Brn67 Homo sapiens cDNA clone IMAGE:4937865

5', mRNA sequence.  
 accession BG910244  
 version BG910244.1 GI:14290720  
 keywords EST.  
 source human.  
 organism Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 reference 1 (bases 1 to 694)  
 authors NIH-MGC <http://mgc.nci.nih.gov/>.  
 title National Institutes of Health, Mammalian Gene Collection (MGC)  
 journal Unpublished (1999)  
 comment Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs@email.nih.gov](mailto:cgabs@email.nih.gov)  
 Tissue Procurement: David N. Louis, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.lnl.gov>  
 Plate: LLM10872 row: f column: 10  
 High quality sequence stop: 694.  
 Location/Qualifiers  
 1..694  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4937865"  
 /clone\_1ib="NCI-CGAP\_Brn67"  
 /tissue\_type="anaplastic oligodendroglioma with 1p/19q  
 loss"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: brain; Vector: PCMV-SPORT6; Site\_1: NCI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 2.3 kb. Constructed by Life  
 Technologies. Note: this is a NCI-CGAP Library."  
 BASE COUNT 186 a 189 c 205 g 114 t  
 ORIGIN  
 alignment\_scores  
 Quality: 1150.00 Length: 227  
 Ratio: 5.134 Gaps: 1  
 Percent Similarity: 98.678 Percent Identity: 97.797  
 alignment\_block:  
 US-09-757-781-2 x BG910244 ..  
 Align seg 1/1 to: BG910244 from: 1 to: 694  
 1094 AspGluuMetTyrGlyValSerSerTyrGlySerMetAlaLeu 1110  
 1 GATGAGTTAATGTAATGGGAGCTTCTTATGAGAGTCCATGAGCTCT 50  
 1110 uAsnAlaArgProGlnSerProArgGlnGlyHisMetAlaLeuT 1127  
 51 CAACGCTAGACCTCAGAGCCAGAGAGGCAATGATGATGCTTGT 100  
 1127 YrIaGlnVallylsYsProArgAsnSerLySProSerProValAspSer 1143  
 101 ATGCCCAAGCAGAGAGCCGCGGAATCCAAACCTCACCCTGTAGACAGT 150  
 1144 AsnArgSerThrProSerAsnHisAspArgTlleglnArgLeuArgGlnG 1160  
 151 AACAGATCAACTCTAGCATCATGATCGATACAGCGTCTGAGCAAGA 200  
 1160 upheGlnGlnAlaLySglnAspGlnAspValGluAspArgArgThr 1177  
 201 ATTTCAGCAAGCAAGCAAGATGAAGATGTAGAAGATCGCGCGGACCT 250  
 1177 YrSerPheGlnGlnProTyrProAsnAlaArgProAlaThrGlnSerGly 1193  
 251 ATAGTTTGACCAACCTGGCGGACGACGCGCGGAGAGAGAGG 300

1194 ArgHisSerValSerValGluValGlnMetGlnArgGlnArgGlnGln 1210  
 301 CGGACCTCGGTGCTCCCTGAGAGTGCAGATGCAGCGCGAGAGAGAG 350  
 1210 uArgGlnSerSerGlnGlnAlaGlnArgGlnTyrSerSerLeuProArg 1227  
 351 GCGCAGAGAGCTCCAGCAGAGGCCAGCCAGCTACAGCTCTGCTCGG 400  
 1227 InSerArgLyAsnAlaSerSerValSerGlnAspSerTyrPheGlnAsn 1243  
 401 AAACAGAGAAAAATGCCAGCTCGCTCCAGGACTCTTGAGAGAGAAC 450  
 1244 TySerProGlnGlnGlyPheGlnInSerAlaLySglnAspProArgTyrSe 1260  
 451 TACCTCCCTGGGAGAGCTTCCAGAGCTGCCAAAGAACCCAGCTACTC 500  
 1260 rSerTyrGlnGlySerArgAsnGlyTyrLeuGlyHisGlyPheAsn 1277  
 501 CACCTACCAAGGCTCCAGAGAGCTACTGAGAGACATGCTTCAACG 550  
 1277 laArgValMetLeu GluThrGlnGlnLeuArgGlnGlnArgGln 1293  
 551 CAGGCTCATGCTGGGAACTCAGAGAGCTCTCGCCAGAGAGAGGCG 600  
 1293 gLySglnGlnGlnMetLySglnProProSerGlnGlyProSerAsnT 1310  
 601 GAAGCAGCAGCAGATGAGAACGACGCTCTCCGAGGGGCCAGCAACT 650  
 1310 YrAspSerTyrTrlySValGlnAspPro 1319  
 651 ATGACTCTATTAAGAAAGTCCAGAGACCA 679  
 seq\_name: gb\_est1:A0079784  
 seq\_documentation\_block:  
 LOCUS A0079784 814 bp mRNA linear EST 12-JUL-2000  
 DEFINITION A0079784 Sugano mouse brain mncb Mus musculus cDNA clone MNCB-4753  
 ACCESSION A0079784  
 VERSION A0079784.1 GI:5084539  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Hashimoto, K., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Toyoda, A.,  
 Suzuki, Y., Sasaki, M. and Sugano, S.  
 Isolation of full-length cDNA clones from a mouse brain cDNA  
 library made by oligo-capping method  
 Unpublished (1999)  
 Contact: Katsuyuki Hashimoto  
 Division of Genetic Resources  
 National Institute of Infectious Diseases  
 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
 Email: [khashi@nih.go.jp](mailto:khashi@nih.go.jp)  
 URL: <http://www.nih.go.jp/yoken/genbank/>.  
 Location/Qualifiers  
 1..814  
 /organism="Mus musculus"  
 /strain="C57BL"  
 /db\_xref="taxon:10090"  
 /clone="MNCB-4753"  
 /clone\_1ib="Sugano mouse brain mncb"  
 /sex="female"  
 /dev\_stage="adult"  
 /lab\_host="TOP10"  
 /note="Organ: brain; Vector: pME18S-Fl3; 1st strand cDNA  
 was primed with an oligo(dT) primer  
 ligated to a DraIII adaptor (TGTGGCCATCTG), digested and  
 cloned into distinct DraIII sites of the pME18S-Fl3. XhoI



sites just outside the DraIII sites can be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5 kb. Library was constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing: 5' end primer [CTTCTCTCTTAAGCTGCG], 3' end primer [CGACGTGACGCTGACACAA].

BASE COUNT 259 a 161 c 246 g 146 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 1146.00 Length: 245  
Ratio: 4.815 Gaps: 0  
Percent Similarity: 97.143 Percent Identity: 88.571

alignment\_block:  
US-09-757-781-2 x AU079784 ..

Align seg 1/1 to: AU079784 from: 1 to: 814

```

855 LeuGlyIleAlaSpGluThrLysLeuAsnThrValAspAspGlnLysAl 871
|||.....|.....|.....|.....|.....|.....|.....|.....|
74 TTAACAGTAGCTGACGAGCTAACTCAATACAGTGGATGACCGAGAGC 123
871 aglySerProSerArgAspValGlyProSerLeuGlyLeuLysLysSers 888
|||||.....|.....|.....|.....|.....|.....|.....|.....|
124 AGGCTCTCCAGTAGAGATGTGGACCCCTTAGTCTCGAACAATCTA 173
888 eSerLeuGluSerLeuGlnThrAlaValAlaGluValThrLeuAsnGly 904
|||||.....|.....|.....|.....|.....|.....|.....|.....|
174 GCTCTTAAAGATCTGCAGACCGCTGTGCCGAGGTGAGCGCTGAACGG 223
905 AspIleProPheHisArgProArgProArgIleIleArgGlyArgGlyCy 921
:::.....|.....|.....|.....|.....|.....|.....|.....|
224 ACATTCCTTCCACCGCCCTCGCCACGAATCATCCGGGAGGGGGCTG 273
921 sAsnGluSerPheArgAlaAlaIleAspLysSerTyrAspLysProAlay 938
|||||.....|.....|.....|.....|.....|.....|.....|.....|
274 CAACGAAACCTTCAGAGCCGCCATTGACAAAGTCTTCGACAAAGCCATG 323
938 aLAspAspAspAspGluGlyMetGluThrLeuGluGluAspThrGlu 954
|||||.....|.....|.....|.....|.....|.....|.....|.....|
324 TTGATGATGACGACGAAAGCATGAGACCTTGGAAAGAACACAGAGGAA 373
955 SerSerArgSerGlyArgGluSerValSerThrAlaSerAspGlnPro 971
|||||.....|.....|.....|.....|.....|.....|.....|.....|
374 ACCTGAGGTCCAGGAGGAGGTCCGTGTCACGTCACGAGTACGCCCTTC 423
971 rHisSerLeuGluArgGlnMetAsnGlyAsnGlnLysGlyAspLys 988
|.....|.....|.....|.....|.....|.....|.....|.....|
424 CTATTCTCTGGAAGAGCAATGAAATGAGACCCAGAAAGAGGACAAGA 473
988 hAspArgLysLysAspLysThrGlyLysGluLysLysLysAspArgAsp 1004
|.....|.....|.....|.....|.....|.....|.....|.....|
474 CAGAGGAGGAAAAAGCAAGCCGGAAAGATTAAGAAAGAACCCGAGAG 523
1005 LysGluLysAspLysMetLysAlaLysLysGlyMetLeuLysGlyLeu 1021
|||||.....|.....|.....|.....|.....|.....|.....|.....|
524 AAGGAGAAAGATTAACCTGAAGCAAGAAAGGGGATCTGAAGGCTTGG 573
1021 yAspMetPheArgPheGlyLysHisArgLysAspAspLysIleGluLys 1038
|||||.....|.....|.....|.....|.....|.....|.....|.....|
574 GGACATGTTCCAGTTTGGCAACATGCAAGAAAGATGACAAGATGAAAAA 623
1038 hGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1054
|.....|.....|.....|.....|.....|.....|.....|.....|
624 TGGGTCGAATTAATAATCCAGATCTTTCACCTCAAAAGAGGACAGGAG 673
1055 ArgMetLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1071
|||||.....|.....|.....|.....|.....|.....|.....|.....|
674 CGGATGAAGAAAGAACAGAGAGAGATTCAAGCAAAACTCGAGAGTTTAA 723

```

```

1071 gGluArgGlnAlaArgGluThrArgAspTyrAlaGluIleGlnAspPheHis 1088
|||||.....|.....|.....|.....|.....|.....|.....|.....|
724 GGAGCGGCAAGCCGAGAGCGTACATATGACAGATTCAGATTTCCATC 773
1088 rGthrPheGlyCysAspAspGluLeuMetLysGly 1099
|.....|.....|.....|.....|.....|.....|.....|.....|
774 GGACGTTTGGCTGTGATGACAAAGTGTCTGTTGGG 808

```

seq\_name: gb\_est1:AL529632

seq\_documentation\_block:

LOCUS AL529632 973 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL529632 LTI\_NFL001\_NBC4 Homo sapiens cDNA clone CS0DD005YD18 5  
prime, mRNA sequence.  
ACCESSION AL529632  
VERSION AL529632.1 GI:12793125  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 973)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Genoscope - Centre National de Sequencage  
Genoscope - BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..973

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CS0DD005YD18"

/clone\_11b="LTI\_NFL001\_NBC4"

/sex="male"

/tissue="neuroblastoma cells"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@life.techn.com URL: http://fulllength.invitrogen.com"

BASE COUNT 209 a 290 c 286 g 171 t 17 others  
ORIGIN

alignment\_scores:  
Quality: 1132.00 Length: 231  
Ratio: 5.122 Gaps: 1  
Percent Similarity: 95.671 Percent Identity: 95.238

alignment\_block:  
US-09-757-781-2 x AL529632 ..

Align seg 1/1 to: AL529632 from: 1 to: 973

```

1 MetLysValThrValCysPheGlyArgThrArgValValAlaProCysG 17
|||||.....|.....|.....|.....|.....|.....|.....|.....|
281 ATGAAGTAAACCGTTRKCTTGGACGACCGGGGTGCTCTKCCGTGCG 330
17 yAspGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34
|.....|.....|.....|.....|.....|.....|.....|.....|
331 GGACGGCCACATGAMAKTTTTCAGCCCTCATCCAGACGCGTGAACCGCT 380
34 yTArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
|.....|.....|.....|.....|.....|.....|.....|.....|
381 ACCGAAAGCCATCGCCAAAGATCCAAACTACTGATACAGGTGATCGC 430

```

```

51 LeuGIUHIISGLYAspGlyIleLeuAspLeuAspAspIleLeuCys 67
|||||
431 TTGGAACTGAGATGAGAGAACTAGACCTTGATACACTTCTTGGA 480
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluAspP 84
|||||
481 TGTAGACAGACGATAAAKACACAGCTGTAGCAGTTTTTGAAGACAGATC 530
84 roHSHISGLYAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
|||||
531 CACATCCAGGAGGTATGACAGCAGTCCAGTCCAGGGGTACCCAGAC 580
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG 117
|||||
581 CCAGAGATATTGTGTAAGTCTGACCCACCAAGTCTCCAGCCCTTCA 630
117 nProTYRGIAlaThrSerGIUleGIUValThrProSerValIleuArg 134
|||||
631 GCCTTACCAAGCAACAAAGTGAATGAGTCAACCTTACGCTTCGAG 680
134 laAsnMetProLeuHisValArgArgSerSerAspProAlaIleuIle 150
|||||
681 CAAATATGCTCTCTCATGTTGCAGCAATAGTTGACCGCTCTAATTGC 730
151 leuSerThrSerValSerAspSerAsnPheserSerGluGluProSerAr 167
|||||
731 CTCTCCACTTCTGTCAGTGTAGTATTTTCTCTGAGAGCCTTCAAG 780
167 gLysAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsn 184
|||||
781 GAAAAATCCACAGCCTGCTCAACAAACGCTGCTCTCAAGCAGCAACA 830
184 hPAla.GlySerProLysThrCysAspArgLysLysAspGluAsnTYR 200
|||||
831 CTGCGGGAGTCTCTAAACCTGCGACAGAAAGATTAACAACTACAG 880
200 gSerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217
|||||
881 AAGCTCCCGGGAGTACTAGTAACTGTGTAACTTCAATTCACAGAGACA 930
217 snAlaArgSerSerLeuSerAlaSerHisProMetValGly 230
|||||
931 ATGCTCGCTGCTCTGAGTGCACAGTCCACCAATGTTGGGC 971
seq_name: gb_est2:B1464081
seq_documentation_block:
LOCUS B1464081 830 bp mRNA linear EST 21-NOV-2001
DEFINITION M63202858F1 NIH_MGC_97 Homo sapiens cDNA IMAGE:5268446 5',
ACCESSION B1464081
VERSION B1464081.1 GI:15254737
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 830)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabps-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraiki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LLM11677 row: d column: 15

```

```

FEATURES             High quality sequence stop: 722.
source               Location/Qualifiers
                    1..830
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5268446"
                    /clone_id="NIH_MGC_97"
                    /lab_host="DH10B"
                    /note="Organ: testis; Vector: pBluescript (modified
                    pBluescript KS+); Site_1: BamHI; Site_2: SalI; XhoI (gtcgag
                    ); Oligo-dT primed using primer 5'-TTTTTTTTTTTAA-3',
                    size-selected for average insert size 2.2 kb and
                    normalized to ROP 5. This is a primary library enriched
                    for full-length clones and constructed using the
                    Cap-trapped method (Carninci, in preparation). Library
                    constructed by M. Brownstein (NHGRI, National
                    Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT          247 a 191 c 204 g 188 t
ORIGIN
alignment_scores:
  Quality: 1127.50      Length: 315
  Ratio: 4.528          Gaps: 7
  Percent Similarity: 79.048  Percent Identity: 74.603
alignment_block:
US-09-757-781-2 x B1464081 ..
Align seg 1/1 to: B1464081 from: 1 to: 830
132 LeuArgAlaAsnMetProLeuHisValArgArgSerSerAspProAla 148
::: |||||
21 ATGACACGCTGATATGCTCTTCAAGTTCGACGACGAGTACCCAGCTCT 70
148 UIleGlyLeuSerThrSerValSerAspSerAsnPheserSerGluGlu 165
|||||
71 AATTGGCCTCCACTCTGTCAGTATGTAATTTTCTCTGAAAGAC 120
165 roSerArgLysAsnProThrArgTrpSerThrThrAlaGlyPheLeuLys 181
|||||
121 CTTCAGAGAAAATCCACACAGCTGTCAACAAACGCTGCTCTCAAG 170
182 GlnAsnThrAlaGlySerProLysThrCysAspArgLysLysAspGluAs 198
|||||
171 CAGAACACTGCTGGAGTCTTAACCTGCCACAGAGAG..... 209
198 nTYRArgSerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGln 215
209 .....
215 rGAspAsnAlaArgSerSerLeuSerAlaSerHisProMetValGlyLys 231
209 .....
232 TrPLeuGluLysGlnGluAspGluAspGlyThrGluLysAspAsnSe 248
210 .....
248 rArgValGluProValGluHisAlaAspThrGlyLeuGluHisIleProa 265
239 TCGGTGTGAACCTGTGGACATGCTGACAGGGTTTGGAGCATATGCCA 288
265 snPheSerLeuAspAspMetValLysLeuValGluValProAsnAspGly 281
289 ACTTTTCTGTGAGATATGTAAAGCTGTAGAACTCCCAACATATGA 338
282 GlyProLeuGlyIleHisValAlaProPheSerAlaArgGlyLysArgTh 298
339 GGGCGCTGTGGAAATCATGATGAGCTTTCAGTGTCTCGAGGGCGGAGAAC 388
298 rLeuGlyLeuLeuValLysArgLeuGluLysGlyLysAlaGluHisG 315
|||||

```

```

389 CCGGGGTTATTAGTAAACGATTGCGAAGAGTGCTAAAGCTGAACATG 438
315 luAsnleuPhearGluAsnAspCysIleValArgIleAsnAspGlyAsp 331
439 AAAATCTTTTTCGTGAGAAATGATTGCATGCTCAGATTAATGATGGGAC 488
332 leuArgAsnArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMet 348
489 CTTGCGAAATAGAGATTGTGAAACAGCAACACATATGTTCCGCAACCAT 538
348 lArgThrProIleIleThrPheHisValValProAlaAlaAsnLysGlu 365
539 GCGTACACCATCATTTGTTCCATGTGTTCTCGCAGCAAAATAAAGAGC 568
365 lntYrGluGlnleuSerGlnSerGluLysAsnAsn..TyrTyrSerSerArg 381
589 AGATATGACACATATCCCAAGAGAGACAAATTTACTATTCAAGCCG 638
381 gPheSerProAspSerGln...TyrIleAspAsnArgSerValAsnSera 397
639 TTTTACCCCTGCACGACGATGATTTGACCACAGGAGTGTACACAGT 668
397 la...GlyLeuHis..ThrValGlnArgAla...ProArgLeuAsnHisP 411
689 GCAGGGCTTCCACAGCGGTGCAGAAAGACACCCCGCATGCAACACCC 738
411 oProGluGlnIleAspSer..HisSerArgLeuProHisSerAlaHisP 427
739 GCGTAGACAGATTGACTCTCTCATCAAGAAATACATCATTAGAGGAAGCC 788
428 SerGlyLysProSerAlaProAlaSerAlaProGln 440
789 CTCGGGAAAACCCCATCGGTCCACTCGGACCCAAAG 827

seq_name: gb_est2:BG336977

seq_documentation_block:
LOCUS BG336977 941 bp mRNA linear EST 27-FEB-2001
DEFINITION 602434429P1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4552423 5',
RNA sequence.
ACCESSION BG336977
VERSION BG336977.1 GI:13143415
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 941)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: csaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILN at:
http://image.llnl.gov
plate: L1CM1248 row: 1 column: 08
High quality sequence stop: 638.
FEATURES
source
1..941
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4552423"
/clone_lib="NIH_MGC_46"
/tissue_type="Telomysarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOT87; Site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:

```

```

GCCACAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the Laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 280 a 228 c 236 g 197 t
ORIGIN
alignment_scores:
Quality: 1118.50 Length: 302
Ratio: 4.285 Gaps: 9
Percent Similarity: 86.424 Percent Identity: 80.795
alignment_block:
us-09-757-781-2 x BG336977 ..
Align seg 1/1 to: BG336977 from: 1 to: 941
218 AlaArgSerSerLeuSerAlaSerHisProMetValGlyTyrPLeuG 234
1 GCTGCGTCTGCTCTGAGTGGCCAGTCCACCAATGCTGGGCAAGTGGTGA 50
234 ulysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgValG 251
51 GAAACAGAACAGAGATGAGATGGAGACAGAAAGAGATTAACAGTGTGTG 100
251 luvProValGlnHisAlaSerPheThrGlyLeuGlnHisIleProAsnPheSer 267
101 AACCTGTGGACATGCTGACACGGGTTTGGAGCATTAACCAACCTTTCT 150
268 leuAspAspMetValLysLeuValGluValProAsnAspGlyGlyProLe 284
151 CTGGATGATATGTTAAAGCTCTGTAGAAAGTCCCAACATGAGGCGCTCT 200
284 uGlyIleHisValValProPheSerAlaArgGlyLysArgThrLeuGlyL 301
201 GGGAAATCATATGATAGTGCCTTCACTGCTCGAGCGGAGAAACCTGGGCT 250
301 leuLeuValLysArgLeuGluGlyGlyLysAlaGlnHisGluAsnleu 317
251 TATTAGTAAACGATTTGAGAAAGGTGTAAAGCTGAACATGAATAATCTT 300
318 PheArgGlu..AsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
301 TTTCGTAGACATGATGATGATGTTAGATTAATGATGGCACCTTCGAA 350
334 snArg..ArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgTh 350
351 ATAGAACGATTTGAAACAGCACACATATATGTTGCCAAGCCATGCGTAC 400
350 rProIleIleThrPheHisValValProAlaAlaAsnLysLuuLntYrG 367
401 ACCCATCATTTGTTCCATGTGTTCTCGCAGCAAAATTAAGACAGATG 450
367 lueGlnleuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSer 383
451 AACGACATATCCCAAGAGAGAGAACAAATTAATTCAGCGGTTTACG 500
384 ProAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuH 400
501 CCGTACAGCCCGGTATTGTGACACAGAGAGTGTGAACAGTGCAGGGCTTCA 550
400 sThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAsp 417
551 CACGTCACAGAGACACCCGACTGAACACACCCCT..GAGGAGATGAGCT 599
417 eHisSerArgLeuProHisSerAlaHis..ProserGlyLysProProse 433
600 CTCGCTCAAGACTACATCAT...AGGCACACCCCTCGGGAATAACAGCATC 646
433 rAlaProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerG 450

```

```
|||||
647 GGCTCCAGCCTCGCACCTCAGCATGTATAGTACGACTGTAAAGACGT 696
450 LYTyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLys 466
697 GGTATATACACCAACAATAATAGCCAGAGTAAATTCACCTTCAGAAA 746
467 GlyThrGlyLeuGlyPheSerIleThrSerArgAspValThrIleGln 483
747 GGGACCGAAGGCTGGGAT...CAACCTACTTCCAGA...ACATGCTACCA 790
483 yGlySerAlaProIleTyrValIysAsnIleLeuProArgGlyAla.... 498
791 AGGTGCTGTCATCGATGCAGAAA...CTCTCCCGCGGTGGGCCAAGAA 837
499 .....AlaIle.GlnAspGlyArgLeuLysAlaGlyAspArgLeuIle 512
838 GGATGCCGATTAAGACGAGGAGGAGACTCCAGCGCGGTCTGTATGCTT 885
```







Alignment scores: Quality: 340.50 Length: 880  
 Ratio: 0.871 Gaps: 35  
 Percent Similarity: 44.432 Percent Identity: 21.023

Alignment block:  
 US-09-757-781-2 x US-09-233-086-2 ..

Align seg 1/1 to: US-09-233-086-2 from: 1 to: 5836

```

124 guilegluValThrProSerValLeuArgAlaAsnMetProLeuHisVal 140
|||||
2743 GAGTATATTCACAGATCCCTCACACATGAGTGTATATCCCTGTGCGCA 2792
140 lArgArgSerAspProAlaLeuIleGlyLeuSerThrSerValSera 157
|||||
2793 CATTCACAGGCCACTCCTGTG.....CCCTGTGTGAATG 2827
157 spSerAsnPheserSerlu.....GluProSerArgLys 168
|||||
2828 AACTTCACCTTGTGTACACAGTGTGCATGATTAATGAACCATCCGACTCT 2877
169 AsnProThrArgTyrSerThrThrAla..... 177
2878 CAAGAGCGACAGAACCGGAGAGCATGTCTATCCACGAGCAGACCCGTA 2927
178 .....GlyPheLeuLysGlnAsnT 184
2928 TGGCTATTCCTCCCTGAAATGATGATGAAGAAATTTGTGTCATGGAGTCC 2977
184 hralArgLysProLysThrCysAspArgLysLysAsp..... 196
2978 TACATCTGTACATCACTGAGAGAAACAGTCAACACAGCAGCATTTGAC 3027
197 .....GluAsnTyrArgSerleuProArgAspThrSerAsnTyrPserAs 211
|||||
3028 GACCTGGAATAATCTTAATTCATTAGCAAAA...ACTAGTCTGATTTAGG 3074
211 nglnPheGlnArgAspAsnAlaArgSerleuSerleuSeraHisProM 228
|||||
3075 CATATCCCGAATGATGTCCAGAGTCCATGCTTCATTTGACCTTCG 3124
228 etValIleGlyTyrPheGlnLysGlnGlnAspGlnAspGlyThrGlu 244
|||||
3125 TTGGGCGTCAA....AGGAGGAGCAAGAAATTTGCCTTATATCAA 3168
245 GluAspAsnSerArgValGluProValGlyHisAlaAspThrGlyLeu.. 260
|||||
3169 CACCAACGACGACAGTATTTTCCAGGCTCAGCATACAGGAATGT 3218
260 ..... 260
3219 GTCTCTAGATATGCCATGATACATGTAGTTACCTGAGAGAGAAG 3268
261 .....GluHisIleProAsnPheserleuAspAspMetValLysLeu 274
|||||
3269 GCGAAGAGAGAACTCCAAATTTTAGCCACTGGGGTCCACCGAAT 3318
275 ValGluVal.....ProAsnAspGlyGlyProLeuGlyIleHisVal 288
|||||
3319 GTTGAAATTTTATAGAACCAATGTCT.....CTTGGGATCAGTAT 3362
288 lValProPheSerAlaArgGlyLysArgPheLeu..... 299
|||||
3363 TGT.....GGTGGCAAACTGTATTAAACGCTTAAGA 3397
300 .....GlyLeuLeuValLysArgLeuGluLysGly 310
|||||
3398 AAGGAGAGAGAGCTTAAGGTATATTCATCAACAGTTTGAACACAGT 3447
311 LysAlaGluHisGlnAsnLeuPheArgGlnAsnAspCysIleValArgI 327
|||||

```





96 rglYthrlInSerProgluIlePheglYSerGluldeuGly..... 109  
2750 GAATTCCTCAAGCAGAGCTGTTAGAGATTTAAATATGGACGACCAATCA 2799  
110 .....ThrasnAsnValSerAlaPheGlnPro 118  
2800 GCAGCTGGCAGTGGCCAGCAGCACACCTTCACAAACTGGCTGCTGCAGCT 2849  
119 TTTGlnAlaThrSerGlulIle.....GluValThrProSerValIleuArg 133  
2850 TTAATAGTTCAAGCTGAGATTTCTGAGAGGCTATCTGCTCAGAGCTGTC 2899  
133 gAlaAsnMetProIleuHis.....ValArgYSerSerAspProAlaL 148  
2900 GCTTACACAGCCATTCGCAAAACACTTCACAAAAGAGAAAGATGACAAAGCT 2949  
148 eu.....IleGlyLeuSerThrSerValSerAlaSerAspSerAsn 159  
2950 CATGGAGAGAAAGCCCTAGAGAGATGAGTAAATCATACATGATCTCACT 2999  
160 PheSerSerGluGluProSerArgYSerAsnProThrArgTTPSerThrTh 176  
3000 CAGCGCTCTCTATCCACATCGGAAAAAT..... 3029  
176 rAlaGlyPheLeuYSerGlnAsnThrAlaGlySerProLYsThrCysAsp 193  
3030 .....GTCAATTGTTAAACATGAGACCCACCACCAACACCGTTGGCAG 3069  
193 rGlyLYsAspArgLYsAsnTYrArgSerLeuProArg...AspThrSerAsn 208  
3070 AGTTGGTGGGAAACCTTCTCACCGATCTCAAGATCTGATGACAGAAATCT 3119  
209 TrpSerAsnGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaLase 225  
3120 TTGGCAGGAGTGACAAAACTAATATTCAAAGTCTGTGGCAGATTAA 3169  
225 rHisProMetValGlyLYsTrpLeuGlnLYsGlnLeuGlnAspGlu 242  
3170 T.....AGAACTCCGAAAGAGAGGA 3189  
242 LYrThGlnLYsAspAsnSerArgValGluProValGlnLYsAlaAspThr 258  
3190 AACATGAATCAGACTCTCATCTCATGTAAGACCCCTGGGCAAGCA..... 3333  
259 GLYLeuGlnHisIleProAsnPheSerLeuAspMet..... 271  
3234 .....TATGCTTAGAGGTGCTACACAAAGATG 3262  
272 .....ValLYsLeuValGluValProA 279  
3263 GAGCATATGATCTTCACCCAGAAAGGAGATCACCTTAGTAAGAACTAAAA 3312  
279 snAspArgLYsLYsProLeuGlyIleHisValAlaProPheSerAlaArg 294  
3313 AAGATGCAAAAGTATGGCTTGGATTCATAATTTATGTTGGGAGAAAGATG 3362  
295 GLYGlyArgThrLeuGlyLeuValLYsArgLeuGlnLYsGlyGly 311  
3363 GGAAGACTGGACCTAGGCATATTATTCAGCTCAGTTGCCCTGGAGAGACC 3412  
311 sAlaGlnHisGlnAsnLeuPheArgLYsAsnArgYsIleValArgIleA 328  
3413 AGCTGACTTCATGATGCTTGAAGCCAGAGACCGTTGATCTGTGA 3462  
328 snAspArgLYsLeuArgAsnArgYrPheGlnGlnAlaGlnHisMetPhe 344  
3463 ATATGTTGAGTCTGGAGGAGTCAAGCCACATGCTGCAATTGAATTTTG 3512  
345 ArgGlnAla..... 347  
3513 CAATAATGCAACCTGAAGATGTGACACTTGTATCTCAGCCAAAAAGAAA 3562

---

348 ...MetArgThrProIleIleTrpPheHisVal..... 357  
3563 GATATCCAAAGAGCTTCTACTCTGCTGATCTCAGATCTCACCATGAGATGAAA 3612  
358 .....ValProAlaAla 361  
3613 ACTACATGAAGAAATCTTCTACATTCAGACAGAGCTATATGATTTCTT 3662  
362 AsnLYsGlnIleuLYr..... 366  
3663 TCCAGATATCACCACTGTACGTGCTACCTTGCAGCAGATCTCGAGAA 3712  
366 ..... 366  
3713 CTCCTTTGGCCGCTGCGGGGCTCGGGAAAGAAAGCTGAGTTCTCAAG 3762  
367 .....GluGlnLeuSerGlnSerGlyLYsAsnAsnTYr 377  
3763 ATTCAGAGACTAGAGTGCACAGCTTGTCTCAAGCGAGTCAATGATTCTT 3812  
378 TYrSerSerArgPhe.....Se 383  
3813 TTTGCCAGCCATTTAGGTGACCAAAACCTGCAGAGATCAGACATGGCAG 3862  
383 rProAspSerGlnTYrIleAspAsnArgSerValAsnSerAlaGlyLeu 400  
3863 CCCCTTCCCCTATCTGTATATATCAAGCCAGGAAAGAGATCTTCACGT 3912  
400 IsthValGln.....ArgAlaProArgLeuAsnHisProProGlu 413  
3913 ATAGATACCAAGCAAAACATAAAAGCCAGGACATTTCTGATGTAAGTAT 3962  
414 GlnIleAsp...SerHisSerArgLeuProHisSerAlaHisProSerGly 429  
3963 TATCAGACGCTGGAGATTCAGACATGAGAGAAAGCCATTCACGACGAG 4012  
429 YLYsProSerAlaProAlaSerAlaProGlnAsnValPheSerThr 446  
4013 TCAGGATCATCAAAACACCAAAAGAGAAATCTTCTTTCAGTGAATACAT 4062  
446 hValSerSerArgLYrAspThr.....LYsLYsIleGly 457  
4063 CCAACAAAGATGAATTTTAAACTTTTCTTCATCAGCTCTCAAGCTCGA 4112  
458 LYsArgLeuAsnIleGlnLeuLYsLYsGlyThrGlnLYsLeuGlyPhe 474  
4113 GATATCTTTGAGGTGAAGCTCTTAAATGATTAACACTTGGGGATAG 4162  
474 rIleThr.....SerArgAspValThrIleGlyLYsSerAlaProI 488  
4163 TGTACAGCGAGCTGTGAATACAGAGTGTACAGACATGCTGGC.....A 4203  
488 leryValLYsAsnIleLeuProArgGlyAlaAlaIleGlnAspLYsArg 504  
4204 TTTATGTGAAGCTGTATTTCGCCAGGAGAGCAGAGACTGATGTTGA 4253  
505 LeuLYsAlaGlyAspArgIleuIleGlyValAsnGlyValAspLeuValGly 521  
4254 ATTTCACAAAGAGTATCGGCTCTTACGTCTCAATGCACTTGTAGTCAAG 4303  
521 YLYsSerGlnGlnValValSerLeuLeuArgSerThrLYsMetGln 538  
4304 AGCCACCATTAAGCAAGCTGTGGAAACATCGAGAAATACA.....G 4344  
538 LYrThValSerLeuLeuValPheArgGlnLYsAlaPheHisProArg 554  
4345 GACAGGTGTTCATCTGTATTAGAAAAGGACAAATCTCAACATCTAAA 4394  
555 Glu.....LeuAsnAlaGluProSerGly 562  
4395 GAACATGTCCCGTTAACCCACAGTGTACCTTTTCAGATCAAGATGCCCA 4444  
562 nMetGlnIlePro.....LYsGluThrLYsAlaGlnAspGlu 575

```

4445 AGGTCAAGCCCAAGAAAAGTGAAGAAAACACTCAGCTCAAGACTACAGA 4494
575 sp1leuValleuthrProAspGlyThrArgGluPheLeuThrPheGluVal 591
4495 GCTTGTTC.....ACTGAGCAAAATACATTGAGGTA 4526
592 ProLeu...AsnAspSerGlySerAlaGlyLeuGlyValSerValysgl 607
4527 AAATTTATTTAAAAATAGCTCAGGTCTAGATTCTGTCGAGAGAGA 4576
607 ysnAspSerGlyValAsnHisAlaAspLeuGlyIlePheValLysSerI 624
4577 TAACTTTAATACCGAGCAAAATTAATCCAGCAATAGAGAGGTTAAAAAG 4626
624 le1leasnGlyGlyAlaAlaSerLysAspGlyArgLeuArgValAsnAsp 640
4627 TCTTCTCTGACAGCCAGCAGCAAGAAAGTGAATAATTGATGTAGAGAGT 4676
641 GlnLeuIleAlaValAsnGlyGlySerLeuLeuGlyLysThrAsnGlnAs 657
4677 GTTATCTTGAAGTAGATGAGCCCTTTGAAAGACTATCTCAGCAGAGA 4726
657 palMetGluThrLeuArgArgSerMetSerThrGluGlyAsnLysArg 674
4727 AGTCATATCTCTCTCAGGGAAGTCTCCAGAA..... 4760
674 lyMetIleGlnLeuIleValAlaArgGlyLeuSerLysCysAsnGluLeu 690
4761 ....GTATTCTTGTCTCTCAGCA..... 4781
691 LysSerProGlySerProGlyProGlyLeuProIleGluThrAlaLe 707
4782 .....CCTCACCCTGGTGTCTACCGGAATATGATCTCGCGT 4819
707 uAspAspArgGluArgGlyLeuSerHisSerLeuTyrSerGlyIleGlu 724
4820 T.....TTGACCCCACTTCAGTCCAGCAGCAAG 4848
724 lyLeuAspGlySerProSerArgAsnAlaAlaLeuSerArgIle..... 738
4849 TACTTCCAAACAGCAGTAAGAAGCTCTTTCAGCCATCATGTGTGAGCA 4898
739 .....MetGlyLysSerGlyLysTyrGluLe 747
4899 AGCAGCAGCTCAGATGAATAATGTAATGTCAAGCAAAAGAACAGTG 4948
747 u...SerProThrValAsnMetProGlnAspAspThrValIleIleGlu 763
4949 CAAGTCCCAATCCAGAGAGACAGATTACAGTACAGCAGTGGAGTGAG 4998
763 spAspArgLeuProValLeuProHisIleSerAspGlnSer...Ser 778
4999 AAGATGACTTAGTACAGCTCCAGCAACATATCAAAATTCGACCTGAGT 5048
779 SerSerSerHisAspAspValGlyPheValThrAlaAspAlaGlyThr 795
5049 TCAGCTTTCATCAGACTCTAAGCAACATGATATCACAGCAGCAGAGTCA 5098
795 palAlaLysAlaAlaIleSer..... 801
5099 TCATYAGAGACCCCAAGAGTCACAGATACCAATTGTACCATGTTTACT 5148
802 .....AspSerAlaAspCys 806
5149 ATCCTCAGAAATATCCCAATTAACAGAGTTTGAGAGACACTATCTTCC 5198
807 SerLeuSerProAspValAspProValLeuAlaPheGlnArgGlyLys 823
5199 CCTCTACCAACCGGATATGGCTCTGGGCAAGATTAATCAACCCCAATCAGA 5248
823 eGlyArgGlnSerMetSer 829
::: ::::: |||

```

```

5249 ATCTGCTTCTCTAGTTCCG 5267
seq_name: /cgn2_6/plodata/2/lna/5A_COMB.seq:US-08-596-291-1
seq_documentation_block:
; Sequence 1, Application US/08596291
; Patent No. 5821075
; GENERAL INFORMATION:
; APPLICANT: GONZ, LEONEL JORGE
; APPLICANT: SARAS, JAN
; APPLICANT: CLAESON-WELSH, LENA
; APPLICANT: HELDIN, CARL-HENRIK
; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596,291
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/115,573
; FILING DATE: 01-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: LC461/7000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; TELEX: 92-1742 EZEKIEL
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8040 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 78..7475
; US-08-596-291-1

alignment_scores:
Quality: 327.50 Length: 927
Ratio: 0.741 Gaps: 34
Percent Similarity: 47.681 Percent Identity: 20.928

alignment_block:
US-09-757-781-2 x US-08-596-291-1 ..
Align seg 1/1 to: US-08-596-291-1 from: 1 to: 8040
||| |||||
80 AspGlnLysAspProHisGlyGlyAspGlyThrSerAlaSerThr 96
2712 GATGCCCAAGATATTT.....GAGAGAGCTTGTGTAGAGCCCT 2749

```



```

4430 TCAGATGCCCAAGTCAAGCCCAAGAAAAAGTGAAGAAAAACAACATCAGG 4479
571 LAGLUSPGLUASPILEVALLEUThrProaspGlyThrArgGluPheLeu 587
4480 TCAAGACTACACGCTTTC.....ACTGAAGAAAT 4511
588 ThrPheGluValProLeu...AsnAspSerGlySerAlaGlyLeuGlyVal 603
4512 ACATTGAGGTAAATATTATTAATAATAGCTAGCTAGATTCAGTTT 4561
603 LserValIlySGlyAsnArgSerIySGluSnhIsAlaAspLeuGlyIleP 620
4562 TTCTCAGAGAAATATTATACCGAGCAATTAATGCCAGCATAGTAA 4611
620 heValIySerIleIleAsnGlyValAlaIleSerIyAspGlyArgLeu 636
4612 GGGTTAAAAAGCTTCTTCGCGACACCGACAGCAAGAAAGTGAATAAT 4661
637 ArgValAsnAspGluLeuIleAlaValAsnGlyGlySerLeuLeuGly 653
4662 GATGTAGGAGATGTATCTGAAGAAATGAGAGCTCTTGAAGAGACT 4711
653 sThrAsnGlnAspAlaMetCyluThrLeuArgArgSerMetSerThrGlu 670
4712 ATCTCAGCAGGAAGTCATATCTGCTCAGGGGAGACTGCTCCAGAA... 4757
670 IyAsnIyArgIyMetIleGluLeuIleValAlaIArgIleSerIyS 686
4758 .....GTAATCTTCTGCTCTGTCAGA..... 4778
687 CysAsnGluLeuIySerProGlySerProGlyProGluLeuProI 703
4779 .....CCTCCACCTGGTGCTGCTACCGGAAT 4804
703 eGIuThrAlaLeuAspAspArgGluArgArgIleSerHisSerLeuTyr 720
4805 TGAATCTGCGCTT.....TTGACCCCACTTCAGT 4833
720 eArgIyIleGluGlyLeuAspGluSerProSerArgAsnAlaIleLeuSer 736
4834 CTCACACACAAGTACTTCCAAACAGCATTAAGACTTCTTCAGCCATCA 4883
737 ArgIle.....MetGlyIuSerG 743
4884 TGTGTGAGCAAAAGCACACGCTCAGATGAATAATGTACAGACAAAG 4933
743 IyIySTyTyrGlnLeu...SerProThrValAsnMetProGlnAspAspThr 759
4934 CAAAAAAGTGCAGTCCCATCCAGAGAGACAGTTCACAGTACAGCA 4983
759 alleIleGluAspAspArgLeuProValLeuProHisLeuSerAsp 775
4984 GTGGAGGTGAGAGATGACTTGTACACGCTCCAGCAAAACATATCAAT 5033
776 GlnSer...SerSerSerHisAspAspValGlyPheValThrAlaAs 791
5034 TCGACCTGGAGTTCAGCTTGTGACAGACTCTAAGCAACATGTATTCACA 5083
791 PalagIyThrTPAlaIyAlaIleSer..... 801
5084 GGCACAGAGTCATCATGAGCACCCCAAGACTCAAGAAATACCATTTGTA 5133
802 .....Asp 802
5134 CCATGTTTACTCTCCTCAGAAAAATCCCAATAAACACAGGTTTGAGGAC 5183
803 SerAlaAspCysSerLeuSerProAspValAspProValLeuAlaPheG 819
5184 AGTAATCTCTCCCTCTACACACCGAGATATGCTGGGAGAGATTATCA 5233
819 natGgluGlyPheGlyArgGlnSerMetSer 829

```

5234 ACCCCAAATCAGAAATCTGCTCTCTACTGCG 5264

seq\_name: /cgn2\_6/plodata/2/lna/6A\_COMB.seq:us-09-100-804-1

seq\_documentation\_block:

Sequence 1, Application US/09100804

Patent No. 6066472

GENERAL INFORMATION:

APPLICANT: GONZ, LEONEL JORGE

APPLICANT: SARAS, JAN

APPLICANT: CLAESON-WELSH, LENA

TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL

TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MASSACHUSETTS

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/100,804

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/596,291

FILING DATE: 09-AUG-1996

APPLICATION NUMBER: US 08/115,573

FILING DATE: 01-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/09943

FILING DATE: 01-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616

REFERENCE/DOCKET NUMBER: 10461/7003

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 8040 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: HOMO SAPIENS

FEATURE:

NAME/KEY: CDS

LOCATION: 78..7475

US-09-100-804-1

alignment\_scores:

Quality: 327.50 Length: 927

Ratio: 0.741 Gaps: 34

Percent Similarity: 47.681 Percent Identity: 20.928

alignment\_block:

US-09-757-781-2 x US-09-100-804-1 ..

Align seg 1/1 to: US-09-100-804-1 from: 1 to: 8040

```
80 AspgluGlnAspProHISGlyAspGlyThrSerAlaSerSerth 96
   ||| |||||
2712 GATGCCCAAGATATTT.....GAGAGACCTTGTTAGAGACCT 2749
   ||| |||||
96 rGlyThrGlnSerProGlnIlePheGlySerGluLeuGly..... 109
   ||| |||||
2750 GAATCTCCAGACAGACTCTGTAGAGATTAATATGGAGCAGCATCA 2799
   ||| |||||
110 .....ThrAsnValSerAlaPheGlnPro 118
   ||| |||||
2800 GCACGTGGCAGCTGGCCAGCAGCACCTCCACAAACTGCTGTGCACCT 2849
   ||| |||||
119 TyrglnAlaThrSerGluIle.....GluValThrProSerValLeuAr 133
   ||| |||||
2850 TTTATGATTCAAGCTGAGATTTCTGAAGAGGCTTCCTGCTCAGACCTGC 2899
   ||| |||||
133 galaSmetProLeuHIS.....ValArgArGSerSerAspProAlaL 148
   ||| |||||
2900 GCTTACCGACCATTTGCCAAACAGTTCAAAAGACAAAGATGACAAAGCTT 2949
   ||| |||||
148 eu.....IleGlyLeuSerThrSerValSerAspSerAsn 159
   ||| |||||
2950 CATTGGAGAGAAAGCCTAGAGAGATGAGTAAATCATACATGATGATCAGT 2999
   ||| |||||
160 PheSerSerGluGluProSerArgLysAsnProThrArgTrpSerThrth 176
   ||| |||||
3000 CAGGCTCTCTCTATCCACATCGCAAAAT..... 3029
   ||| |||||
176 rAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysThrCyAspA 193
   ||| |||||
3030 .....GTCAATTGTTAACATGAGAACCCCAACCAACCGTTGCG 3069
   ||| |||||
193 rGlyLysAspGluAsnLysArgSerLeuProArg...AspThrSerAsn 208
   ||| |||||
3070 AGTTGGTGGAAACCTCTCCACAGATGTCAAGATCTGATGCAAACT 3119
   ||| |||||
209 TrpSerAsnGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSe 225
   ||| |||||
3120 TTGGCAGAGAGTGCACAAACTTAATATCAAAAGTCTGTGGCAATTTAA 3169
   ||| |||||
225 rHISProMetValGlyLysTrpLeuGluLysGlnGluGlnAspG 242
   ||| |||||
3170 T.....AGAGTCTGAAAGAGAGCA 3189
   ||| |||||
242 LysThrGluGlnAspAsnSerArgValGluProValGlyHISAlaSerThr 258
   ||| |||||
3190 AACATGAATCAGACTCTCATCATGATGAGACCTGGGCAAGCA..... 3233
   ||| |||||
259 GlyLeuGlnHISLeuProAsnPheSerLeuAspAspMet..... 271
   ||| |||||
3234 .....TATGTTCTAGATGCTACACAAAAGATG 3262
   ||| |||||
272 .....ValLysLeuValGluValProA 279
   ||| |||||
3263 GAGCATGATATCTTACACAGAAAGAGATCTACTGAACTGAA 3312
   ||| |||||
279 snAspGlyGly...ProLeuGlyIleHISValValProPheSerAlaArg 294
   ||| |||||
3313 AAGATGCAAAGTATGCTTGGATTTCAATTT..... 3347
   ||| |||||
295 GlyGlyArg.....ThrLeuGlyLeuLeuValLysArgLeuG 307
   ||| |||||
3348 GGTGGGGAAGAGATGAGACTGACCTAGGCATTTATATCAGCTCAGTTGC 3397
   ||| |||||
307 LysGlyLysLysAlaGlnHISGlnAsnLeuPheArgGluAsnAspCysI 324
   ||| |||||
3398 CCGTGGAGAGCAGCTGATCTTCATGATGCTTGAAGCCAGAGACCTTT 3447
   ||| |||||
324 LevalArgGlyLeuAspGlyAspLeuArgAsnArgArgPheGlnGlnAla 340
   ||| |||||
3448 TGAAATCTGTGAATAGTGTGAGTGGAGGAGTACGACCACTGCTGCA 3497
```

```
341 GlnHISMetPheArgGlnAla..... 347
   ||| |||||
3498 ATTGAATTTTGCAGAAATGCAACCTGAAGATGTCACACTTTTATCTGCA 3547
   ||| |||||
348 .....MetArgThrProIleIleTrpPheHISVal.... 357
   ||| |||||
3548 GCCAAAGAGAAAGATATCCAAAGTGCCTTACTCTGTGCATTCACCA 3597
   ||| |||||
357 ..... 357
   ||| |||||
3598 ATGAGATGAAAAACTACATGACAAATCTCTACATGACAGACAGTCT 3647
   ||| |||||
358 ValProAlaAlaAsnLysGlnGlyArg..... 366
   ||| |||||
3648 ATAGATTCTTCTTCCAGAGATCACACCTGTGCACCTGTACCTGAGGCA 3697
   ||| |||||
366 ..... 366
   ||| |||||
3698 CATTGGAGAACTCTTGGGCGCTGTGGGCGCTGGGAGAGAAAGCC 3747
   ||| |||||
367 .....GluGlnLeuSerGlnSerGlu 373
   ||| |||||
3748 TGAGTCTCAAGATTCAGAGCTGAGAGTGCAGCTGTCTCAAGCCAG 3797
   ||| |||||
374 LysAsnAsnLysTrpSerSerArgPhe..... 382
   ||| |||||
3798 GTCAATGGTTCCTTTCGACCCATTAGGTGACCAACCTGGCAGAAATC 3847
   ||| |||||
383 .....SerProAspSerGlnTrpIleAspAsnArgSerValAsn 396
   ||| |||||
3848 ACACATGACGACGCCCTCCCATCTGTATATCCAAACCCAGAGAAAG 3897
   ||| |||||
396 eAlaGlyLeuHISThrValGln.....ArgAlaProArgLeuAsn 409
   ||| |||||
3898 AGACTTTCAGTATGATAACCAACCAAACTTAAAGCCAGCACTTCT 3947
   ||| |||||
410 HISProGlnGlnIleAsp...SerHISerArgLeuProHISerAl 425
   ||| |||||
3948 GATGTAAGTATGATCTCAGACCTGAGATTCAGACATGATGAGAGCCAC 3997
   ||| |||||
425 aHISProSerGlyLysProProSerAlaProAlaSerAlaProGlnAsn 442
   ||| |||||
3998 TTACTCCAGCAGTCAAGATCATCAACACCAACCAAGCAATCTCTCT 4047
   ||| |||||
442 aPheSerThrThrValSerSerGlyTrpAsnThr..... 453
   ||| |||||
4048 CAGTGAATATCATCCAAACAAAGATGATTTAAACTTTTCTTCATCACT 4097
   ||| |||||
454 LysLysLleGlyLysArgLeuAsnIleGlnLeuLysGlyThrGluG 470
   ||| |||||
4098 CCTAAGCTGGAGATATCTTGTAGCTTCACTGGCTAAATATATAACAG 4147
   ||| |||||
470 LysGlyPheSerIleThr.....SerArgAspValThrIleGluG 484
   ||| |||||
4148 CTTGGGAGATTAAGTGTACAGGAGGTGTGAATACGAGTGCAGCATGGTG 4197
   ||| |||||
484 LysSerAlaProIleTyValLysAsnIleLeuProArgGlyAlaIle 500
   ||| |||||
4198 GC.....ATTATGTGAAGATGTTATTTCCCAAGGAGACACAGAG 4238
   ||| |||||
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValLysGly 517
   ||| |||||
4239 TCTGATGTAGAAATTCACAAAGTGTATCGCTCAGCTGCAATGAGGT 4288
   ||| |||||
517 LAspLeuValGlyLysSerGlnGlnGluValValSerLeuLeuArgSer 534
   ||| |||||
4289 TAGTCTGAAGAGAGCCCAATACCAAGCTGTGGAACCACTAGAGAAATA 4338
   ||| |||||
534 hTrpMetGlnGlyThrValSerLeuLeuValPheArgGlnGlnAspAla 550
   ||| |||||
4339 CA.....GGACAGTGGTTCATCTGTTATTTAGAAAGGAGCAATCT 4379
```

```

551 PheHisProArgLuu.....LeuAsnAl 558
      :::|||
4380 CCACATCTTAAGAACATGTCCGGTAACCCACAGTGTACCTTCACG 4429
558 agLuproSerGlnMetGlnLlePro.....LysGluThrLysA 571
      ::: :::||| ||| |||
4430 TCAGAAATGCCCAAGCTCAAGAGCCCAAGAAAGTGAAGAAACAACTCAGG 4479
571 LgLUspLusApLIleValLleuThrProAspGlyThrArgGluPheLeu 587
      ::::||||| ::::|||||
4480 TCAAGACCTACAGCTTGTG.....ACTGAAGAAAT 4511
588 ThrPheGluValProLeu..AsnAspSerGlySerAlaGlyLeuGlyA 603
      ||||| ||| ::::||||| ::::|||||
4512 ACATTTGAGGTAAATATTATTAAATATACCTCAGCTGTAGATTACGTT 4561
603 LserValLysGlyAsnArgSerLysGluAsnHisAlaAspLeuGlyLle 620
      ||| ::::||||| ::::|||||
4562 TTCTCGAGAAATATCTTATACCGGAGCAAAATTAAATGCCAGCATAGTA 4611
620 heValLysSerLleIleAsnGlyLAlaAlaSerLysAspGlyArgLeu 636
      ||||| ::::||||| ::::|||||
4612 GGGTTAAAAAGCTTGTGTCGACACGACGACGAAAGTGAAGAAAT 4661
637 ArgValAsnAspGlnLeuIleAlaValAsnGlyLusSerLeuGlyLys 653
      ||||| ::::||||| ::::|||||
4662 GATGTGGAGATGTATCTTGAAGATGAGCTCTTTGAAGAGACT 4711
653 snrHisnGlnAspAlaMetGluThrLeuArgArgSerMetSerThrGluG 670
4712 ATCTCGACGAGAGATCATCTGCTCTCAGGGAGACTGCTCCAGAA... 4757
670 LysnLysArgLysMetLleGlnLeuIleValAlaArgArgLysSerLys 686
4758 .....GTATCTTGTCTCTCTCAGAA..... 4778
687 CysAsnGlnLeuLysSerProGlySerProGlyProGluLeuProIle 703
4779 .....CCTCCACCTGTGTGCTACCGGAAAT 4804
703 eglThrAlaLeuAspAspArgLysArgLysSerHisSerLeuTyrS 720
      ||||| |||
4805 TGTACTGTGCTT.....TTGACCCCACTTCAG 4833
720 ergLylLleGlyLeuAspGlySerProSerArgAsnAlaAlaLeuSer 736
      || ::::||||| ::::|||||
4834 CTCACGACAAAGACTTCCAAACAGAGTAAGACTCTTCTGCACATCA 4883
737 ArgLle.....MetGlyLusSergI 743
      :::
4884 TGTGTGGAGCAAAAGCACAGCTCAGATGAATAATGTACAGACAAAG 4933
743 LysTyrGlnLeu..SerProThrValAsnMetProGlnAspAspThyV 759
      :||| ||| ::::||||| ::::|||||
4934 CAAAAAACGTCAAGTCCCATCCAGAAAGACAGTTCACAGTCACAGCA 4983
759 alIleLleGlnAspAspArgLeuProValLeuProHisLysLeuSerAsp 775
      ::::||||| ::::|||||
4984 GTGGAGGTGAGAGATGACTTAGTCACAGCTCCAGCAAAACATATCAAT 5033
776 GlnSer..SerSerSerHisAspAspValGlyPheValThrAlaAs 791
      ::::||||| ::::|||||
5034 TCGACCTGAGTACGCTTGTGCTCAGACTCTAAGCAACATGCTATCACA 5083
791 palagLysThrPalalysAlaAlaIleSer..... 801
      :||| ::::||||| ::::|||||
5084 GGCACAGAGTCATCATGAAGACCCCAAGAGTCAAGAAATACCATTTGTA 5133
802 .....Asp 802
      |||
5134 CCATGTTTACTATCTCTCAGAAATTCACAAATAAACAGAGTTTGAGAC 5183
803 SerAlaAspCysSerLeuSerProAspValAspProValLeuAlaPheG 819

```

```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-290-640-45
seq documentation block:
; Sequence 45, Application US/09290640
; Patent No. 6204055
; GENERAL INFORMATION:
; APPLICANT: Dean, Nicholas M.
; APPLICANT: Marcussen, Eric G.
; TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling
; FILE REFERENCE: ISPH-0051
; CURRENT APPLICATION NUMBER: US/09/290,640
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 8119
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64)..(7521)
; PUBLICATION INFORMATION:
; JOURNAL: FEBS Lett.
; VOLUME: 337
; ISSUE: 2
; PAGES: 200-206
; DATE: 1994-01-10
; DATABASE ACCESSION NUMBER: D21209/Genbank
; DATABASE ENTRY DATE: 1999-02-05
; US-09-290-640-45

alignment_scores:
      Quality: 324.50      Length: 933
      Ratio: 0.729      Gaps: 32
      Percent Similarity: 47.696      Percent Identity: 20.686

alignment_block:
US-09-757-781-2 x US-09-290-640-45 ..
Align seg 1/1 to: US-09-290-640-45 from: 1 to: 8119

80 AspGluGlnAspProHisHisGlyLysGlyAspGlyThrSerAlaSerSerTh 96
      ||| |||||
2698 GATGCCCAAGATAT.....GAGAGAGCTCGTTTGAAGACCT 2735
96 rgLysThrGlnSerProGluLlePheGlySerGluLeuGly..... 109
      ::::||||| ::::|||||
2736 GAATCTCCAGCAGAGCTGTGTAGAGGATTTAAATATGGCAGCAATCA 2785
110 .....ThrAsnAsnValSerAlaPheGlnPro 118
      ||| |||
2786 GCACCTGCACTGTGGCCAGCAGCACCTCAACAACTTGCTGTTCGACCT 2835
119 TyrGlnAlaThrSerGluLle.....GluValThrProSerValLeuAr 133
      ::::||||| ::::|||||
2836 TTATTCAGTTCAAGCTGAGATTCTGAAGAGCTATCTCTCAGAGCTGTC 2885
133 galAsnMetProLeuHis.....ValArgArgSerSerAspProAla 148
      :||| ::::||||| ::::|||||
2886 GCTTTACACACCATTCGCAAAACAGTTCAAAAGAGAAAGTACAAAGCTT 2935
148 eu.....IleGlyLeuSerThrSerValSerAspSerAsn 159
      :||| |||
2936 CATGGAGAGAAAGCCCTAGAGAGATGAGTAAATCATACATGATCTCAGT 2985

```

160 PheSerSerGluGluProSerArgLysAsnProThrArgTrpSerThrTh 176  
 : : : : :  
 2986 CAGGCTCTCTCTATCCACATCGGAAAT..... 3015  
 176 rAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysThrCysAspA 193  
 : : : : :  
 3016 .....GTCATGTGTACATGTGAMACCCACCACCAACACCGTTGAG 3055  
 193 rGlyLysAspGluAsnTrpArgSerLeuProArg...AspThrSerAsn 208  
 : : : : :  
 3056 AGTTGGTGGGAAACCTTCTCACAGATGTCAAGATCTGATGCGAAATCT 3105  
 209 TrpSerAsnGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSe 225  
 : : : : :  
 3106 TTGGCAGGAGTGCAAAACTTAATTAATTCAAAGTCTGTGCGAGTTTAA 3155  
 225 rHisProMetValGlyLysTrpLeuGluLysGlnGluGlnAspGluAsp 242  
 : : : : :  
 3156 T.....AGAACTCTGMAAAGANGA 3175  
 242 LyrThrGluLysAspAsnSerArgValGluProValGlyHisAla..... 256  
 : : : : :  
 3176 AACATGATCAGACTCTCATCCATGGAAGACCCCTGGCAACCAATATGTT 3225  
 257 .....AspThrGlyLeuGlnHisIleProAs 265  
 3226 CTAGGAATGACTATGCATAGTACTTGGAAACTCTTCATCCCAAGTACCTT 3275  
 265 nPheSerLeuAspAspMet..... 271  
 3276 AAAAGAAATGATGTCTACACAAAGATGAGCATAGTATCTTCACCAG 3325  
 272 .....ValLysLeuValGluValProAsnAspGlyGly...ProLeu 284  
 : : : : :  
 3326 AAAGGAGATCACCCTTAGTGAACCTGAAAAAGATGCAAGTATGCGTTG 3375  
 285 GlyIleHisValValProPheSerAlaArgGlyLysArgThrLeuGlyLe 301  
 : : : : :  
 3376 GGAATTCAAATATATGTGGGAGAAAGATGGAAAGTGGAACTGGACCTAG 3425  
 301 uLeuValLysArgLeuGluLysGlyGlyLysGlnHisGluAsnLeuP 318  
 : : : : :  
 3426 ATTATTCAGTTCAGTTGCCCTGGAGGACCAAGCTGATGGATGATGCT 3475  
 318 heaArgGluAsnAspCysIleValArgIleAsnAspGlyLysPleuArgAsn 334  
 : : : : :  
 3476 TGAAGCCAGAGACCGTTTGATATCTGTGAATAGTGTGAGTGGAGGA 3525  
 335 ArgArgPheGluGlnAlaGlnHisMetPheArgGlnAla..... 347  
 3526 GTCAGCCACCATGTGCATTAATTTTGGCAAAATGCAACCTGAAGATGT 3575  
 348 .....MetArgThrProIle 353  
 3576 GAACACTGTTATCTCTCAGCCAAAGAAAGAAAGATATCCAAAGTGCCTT 3625  
 353 LeTrpPheHisVal..... 357  
 3626 CTCCTGTGCATCTCACAATGATGAAAACTACATGAAGAAATCTTCC 3675  
 358 .....ValProAlaAlaAsnLysGluGlnLys 366  
 3676 TACATGCACAGACAGTGTATGATTTCTTCCAGAGATCACCACTGGTTC 3725  
 366 ..... 366  
 3726 ACGTGTACCTGAGGCACATCTCGAGAACTCTTTGGGCATCTGGGG 3775  
 367 .....Glu 367  
 3776 GCCGCGCGGAGAGAACCTGAGTTCTCAAGATTCAGAGACTAGAGTCCC 3825  
 368 GlnLeuSerGlnSerGluLysAsnAsnTrpLysSerArgPhe..... 382

3826 AGCTGTCTCAAAAGCAGGATGATGTTCTTTGGCCAGCATTTAGTGA 3875  
 : : : : :  
 383 .....SerProAspSerGlnTrpIleA 390  
 3876 CCAAACTGGCAGGAAATCACAGCATGGACCCCTTCCCATGTGAATAT 3925  
 390 sPAsnArgSerValAsnSerAlaGlyLeuHisThrValGln..... 403  
 : : : : :  
 3926 CCAAAAGCCAGGAGAAAGACTTCTACATGATGATACCAAGCAAAACT 3975  
 404 ArgAlaProArgLeuAsnHisProProGluGlnIleAsp...SerHisSe 419  
 : : : : :  
 3976 AAAAGCCAGGCAATTTCTGATGTAACTGATTAATCTACAGACCGTGA 4025  
 419 rArgLeuProHisSerAlaHisProSerGlyLysProProSerAlaPro 436  
 : : : : :  
 4026 AGACATGATGAAAGCCACTTCTCCAGCATGACAGATCATCAACACCA 4075  
 436 LaSerAlaProGlnAsnValPheSerThrValSerSerGlyTrpAsn 452  
 : : : : :  
 4076 AACAGGAATCTTCTCTCACTGATACATCAACCAAGATGATTTTAA 4125  
 453 Thr.....LysLysIleGlyLysArgLeuAsnIleGln 464  
 : : : : :  
 4126 ACTTTTCTTCATCACCCTCAAGCCGAGATATCTTTGAGGTGAACT 4175  
 464 uLysLysGlyThrGluGlyLeuGlyPheSerIleThr.....Sera 478  
 : : : : :  
 4176 GGTAAAAATGATTAACAGCTTGGGATAGTGCACGGAGGTGATTA 4225  
 478 rGAspValIleThrIleGlySerAlaProIleTrpValLysAsnIleLeu 494  
 : : : : :  
 4226 CGAGTGCACACATGCTGGC.....ATTATGTCAAAAGCTGTAA 4266  
 495 ProArgGlyAlaIleAlaIleGlnAspGlyArgLeuLysAlaGlyAspArg 511  
 : : : : :  
 4267 CCCAGGAGGACAGCAGATGCTGATGTAGAAATTCACAAAGGTGATCCGT 4316  
 511 uIleGluValAsnGlyValAspLeuValGlyLysSerGlnGluValVal 528  
 : : : : :  
 4317 CCTAGCTCAATGAGATGATGCTAGAAAGGACCAACCATTAAGCAAGCT 4366  
 528 aSerLeuLeuArgSerThrLysMetGluGlyThrValSerLeuLeuVal 544  
 : : : : :  
 4367 TGAAGAACACTGGAATACA.....GACAGGTGCTTCATCTGTTA 4407  
 545 PheArgGlnGluAspAlaPheHisProArgLys..... 555  
 4408 TTGAAGAAAGGACATCTCCACATCTMAAGAACATGTCCCGTAAACCC 4457  
 556 .....LeuAsnAlaGluProSerGlnMetGlnIlePro..... 566  
 4458 ACAGTGTACCTTTCAAGATCAGAAATCCCAAGGTCAAGGCCCAAGAAAG 4507  
 567 .....LysGluThrLysAlaGluAspGluAspIleValLeuThrProAsp 581  
 : : : : :  
 4508 TGAAGAAACAACTCAGTCAAGCAAGACTACAGCTTGTCTC..... 4545  
 582 GlyThrArgGluPheLeuThrPheGluValProLeu...AsnAspSerG 597  
 : : : : :  
 4546 .....ACTGAAGAAATATACATTTAGGTAATAATTAATAATACTC 4589  
 597 ySerAlaGlyLeuGlyValSerValLysGlyAsnArgSerLysGluAsnH 614  
 : : : : :  
 4590 AGCTTAGGATTCAGTTTCTCGAGAAAGATATCTTAACCGAGACAAA 4639  
 614 IsAlaAspLeuGlyIlePheValLysSerIleIleAsnGlyAlaAla 630  
 : : : : :  
 4640 TTAATGCCAGCATAGTAAAGGTTAAAAAGCTCTTCTCGACAGCCAGCA 4689  
 631 SerLysAspGlyArgLeuArgValAsnAspGlnLeuIleAlaValAsnG 647  
 : : : : :



```

4690 GCAGAAAGTGGAAAAATGATGATAGAGATGTTATCTTGAAGTGAATGG 4739
647 yGluSerLeuLeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArg 664
      ||||| ||| :||||:||||:||||:||||:||||:
4740 AGCCTTTTGAAGACTATCTCAGCAGAGATCATATCGCTCTCAGG 4789
664 rGSerMetSerThrGluGlyAsnLysArgGlyMetIleGlnLeuIleVal 680
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4790 GAACCTCTCCAGAA.....GTATCTTGCTTCCTC 4818
681 AlaArgArgIleSerLysCysAsnGluLeuLysSerProGlySerProP 697
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4819 TGCAGA.....CCTCCACC 4832
697 oglyProGluLeuProIleGluThrAlaLeuAspAspArgGluArgArg 714
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4833 TGGTGTGCTACCGGAATTCATCTCGGCT..... 4863
714 leSerIleSerLeuTySerGlyIleGluGlyLeuAspGlySerProSer 730
      ||| ||| ||| :||| :||| :||| :||| :||| :||| :|||
4864 ..TTGACCCCACTTCAGTCTCCAGCACAAAGTCTCCAAACACGAGTAA 4911
731 ArgAsnAlaIleLeuSerArgIle..... 738
4912 GACTCTTTCAGCATCATGTGTGGACAAAGCACACGCTCAGATGAAA 4961
739 ...MetGlyLysSerGlyLysTyGlnLeu...SerProThrValAsnM 753
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4962 TGAATGTCCAGCAAAAGCAAAAGCAAGTCACCTCCCAACGAAAG 5011
753 eLProGlnAspAspThrValIleIleGlnAspAspArgLeuProValLeu 769
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5012 ACAGTTACAGTGCACAGTGGAGTGGAGAGATGACTTACGACAGCT 5061
770 ProProHisLeuSerAspGlnSer...SerSerSerSerHisAspAsp 785
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5062 CCAGCAACATCATCATTCGACCTGAGTTCAGCTTGCATCAGACTCT 5111
785 lGlyPheValThrAlaAspAlaGlyThrThrAlaLysAlaIleSer. 801
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5112 AAGCAACATGTTATCAGCAGCAGAGTCAATGAAAGCAACCAAGCTC 5161
801 ..... 801
5162 AAGAGATACCAATTTGTACCATGTTTACTATCCAGAAATTTCCCAAT 5211
802 .....AspSerAlaAspCysSerLeuSerProAspValAs 813
5212 AAACGAGATTGGAGACAGTAAATCTCCCTCTACCCAGCGATATGGC 5261
813 pProValIleuAlaPheGlnArgGluGlyPheGlyArgGlnSerMetSer 829
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5262 TCCTGGCAGAGTTATCAACCCCAATCAGAAATCTGCTCTAGTTCCG 5310
seq_name: /cgn2_6/plodata/2/1na/5A_COMB.seq:US-08-410-804-2
seq_documentation_block:
; Sequence 2, Application US/08410804
; Patent No. 5632994
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Sato, Takaki
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive, Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,804
; FILING DATE: 27-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/259,514
; FILING DATE: 14-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1389
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-410-804-2

```

```

alignment_scores:
  Quality: 275.50      Length: 533
  Ratio: 1.036        Gaps: 19
  Percent Similarity: 49.906   Percent Identity: 24.765

```

```

alignment_block:
US-09-757-781-2 x US-08-410-804-2 ..

```

```

Align seq 1/1 to: US-08-410-804-2 from: 1 to: 1830

```

```

360 AlAlaAsnLysGlnGlnTyGlnLeuSerGlnSerGlnLysAsn.. 375
      |||:||||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
34 GCCACCGAAGAGACTTTCAGTCAATGATGATACCAAGCAAAAGTAAATA 83
376 .....AsnTyTySerSerArgPheSerProAsp 386
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
84 GCCAGCATTTCTGATGATGATCTACTCAGACCGTGGAGATTCAGAC. 132
386 eGlnTyIleAspAsnArgSerValAsnSerIleGlnLeuHisThrVal 402
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
133 .....ATGATGAAGCCATTACTCAGACATCAGATCAT..... 168
403 GlnArgAlaProArgLeuAsnHisProGlnGlnIleAspSerHisSe 419
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
169 .....CAACACCAAAACAG...GAATCTTCCTC 194
419 rArgLeuProHisSerAlaHisProSerGlyLysProProSerAlaProA 436
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
195 TTCAGTGAATATACATCAACAGATGAATTTTAAACTTTTCTTCACAC 244
436 laSerAlaProGlnAsnValPheSerThrThrValSerSerGlyTyAsn 452
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
245 CTCCTAAGCCTGGAGATATCTTT..... 267
453 ThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysGlyThrG1 469
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
268 .....GAGGTGAAGTGGCTTAATAATGATGATA 293
469 uGlyLeuGlyPheSerIleThr.....Sera 478
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
294 CAGCTTGGGATAGTGTACAGGACTGTTTGACAGGAGAGTGTGATA 343
478 rGaSPValThrIleGlyLysSerAlaProIleTyValLysAsnIleLeu 494
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
344 CGAGTGCAGACATGATGCC.....ATTATGTGAAAGCTGTATT 384

```

```

495 ProArgGlyAlaAlaIleGlnAspGlyArgLeuYsaIaGlyAspArgLeu 511
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
385 CCCAGGGAGCACAGAGTCGTAGTATTCACAAAGGATGATCGGCT 434
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
511 uIleGlyAlaAsnGlyValAspLeuValGlyLysSerGlnGluValY 528
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435 CCGAGCTGCATGAGTATGCTAGAAAGAGCCACCCATAGCAAGCTG 484
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
528 aISerLeuArgSerThrLysMetGluGlyThrValSerLeuVal 544
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
485 TGGAAACACTGAGAAATACA.....GGACAGTGTTTCATCTGTTA 525
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
545 PheArgGlnGluAspAlaPheHisProArgGlu..... 555
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
526 TTGAAAGAGGCAATCTCCAAATCTTAAGAACATGTCGCCGTAAACCC 575
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
576 ACAGTGTACCTTTGATCAGATCAGAAATGCCCAAGGTCAGAGCAAAAG 625
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
567 ....LyGluThrLysAlaGluAspGluAspIleValLeuThrProAsp 581
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
626 TGAAGAAACAACTCAGCTCAAGACTACAGCTTGTG..... 663
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
582 GlyThrArgGluPheLeuThrPheGluValProLeu...AsnAspSerG 597
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
664 .....ACTGAAGAAATACATTGAGGTAAATATTATTAATAATACCTC 707
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
597 ySerAlaGlyLeuGlyValSerValLysGlyAsnArgSerLysGluAsn 614
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
708 AGGCTTAGAGATTCTCTCTCAGAAAGAACTTAATACCGAGAGCAAA 757
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
614 ISaIaAspLeuGlyIlePheValLysSerIleLeuAsnGlyAlaAla 630
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
758 TTAATGCCAGCATAGTAAGGTAAAGCTCTTCTCGAGCAGCAGCA 807
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
631 SerLysAspGlyArgLeuArgValAsnAspGlnLeuIleAlaValAsn 647
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
808 GCAGAAAGTGAAATATGATGATGAGGATGATATCTTGAAGTGAATG 857
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
647 yGluSerLeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArg 664
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
858 AGCCTCTTTGAAGACTATCTCAGCAGGAGACATATCTGCTTCAGCG 907
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
664 ySerMetSerThrGluGlyAsnLysArgGlyMetIleGlnLeuVal 680
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
908 GAAGCTCTCCAGAA.....GATTCCTGCTCTC 936
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
681 AlaArgArgIleSerLysCysAsnGluLeuLysSerProGlySerPro 937
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
937 TGCAGA.....CCTCCACG 950
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
697 cGlyProGluLeuProIleGluThrAlaLeuAspArgGluArgArg 714
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
951 TGGTGTGCTACCGGAATGTGATATCGCGTT..... 981
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
714 leSerHisSerLeuYSerGlyIleGluGlyLeuAspGluSerProSer 730
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
982 ..TTGACCCCACTTCAGTCTCCAGCACAAGTACTTCAAAACAGCAGTAA 1029
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
731 ArgAsnAlaAlaLeuSerArgIle..... 738
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1030 GACTCTCTCAGCCATCATGTGTGAGCAAAAGCAGCAGTCAATGAAA 1079
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
739 ...MetGlyGluSerGlyLysTyrGlnLeu...SerProThrValAsn 753
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1080 TGAATGTGACAGCAAAAGCAAAAGAGTGCAGATCCCAATCCAGAGAG 1129
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
753 etProGlnAspAspThrValIleIleGlnAspAspArgLeuProValLeu 769
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1130 ACAGTTACAGTACAGCAGCTGGAGTGAGAAAGATGACTTAGTGACGCT 1179
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
770 ProProHisLeuSerAspGlnSer...SerSerSerHisAspAspVa 785

```

```

      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1180 CCAGCAACAATCAATTCAGCTGAGTTCAGCTTTCATCAGACTCT 1229
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
785 IgyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIleSer 801
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1230 AAGCAACATGTATACAGCAGCAGAGTATCATGAAAGCAGCAGAGTC 1279
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
801 ..... 801
1280 AAGAATACCATTTTACCATGTTTACTATCTCAGAAATTCACAT 1329
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
802 .....AspSerAlaAspCysSerLeuSerProAspValAs 813
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1330 AAACCAAGTGTGAGCAGACAGATATCTTCCTCCCTACACACCGATATGC 1379
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
813 pProValLeuAlaPheGlnArgGluGlyPheGlyLysArgLysSerMetSer 829
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1380 TCCTGGGCGAGTTATCAACCCCAATCAGAAATCTGCTCTGTAGTTG 1428
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/prodata/2/ina/5A_COMB.seq:us-08-259-514-2

```

```

seq_documentation_block:
; Sequence 2, Application US/08259514
; Patent No. 5747245
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive, Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/259,514
; FILING DATE: 14-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9954
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-259-514-2

```

```

alignment_scores:
      Quality: 275.50      Length: 533
      Ratio: 1.036      Gaps: 19
      Percent Similarity: 49.906      Percent Identity: 24.765

```

alignment\_block:

```

US-09-757-781-2 x US-08-259-514-2 ..
Align seg 1/1 to: US-08-259-514-2 from: 1 to: 1830

```

360 AlaAlaAsnLysGluGlnTyrGluGlnLeuSerGlnSerGlnLysAsn.. 375



```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,311
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/410,804
FILING DATE: 27-MAR-1995
APPLICATION NUMBER: US 08/259,514
FILING DATE: 14-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1389
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-858-311-2

alignment_scores:
Quality: 275.50 Length: 533
Ratio: 1.036 Gaps: 19
Percent Similarity: 49.906 Percent Identity: 24.765

alignment_block:
US-09-757-781-2 x US-08-858-311-2 ..

Align seg 1/1 to: US-08-858-311-2 from: 1 to: 1830
360 AlaAlaasnlysglnlytgrglleuSerGlnSerGlnlyulysasn.. 375
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 GCCACCGAAGAAAGACTTTCACGTAGTACCAAAAGCAAACTAAAA 83
376 .....AsnlytYrSerSerAlaGpHeSerProAspS 386
84 GCCAGCATTTTCGTACTACTATCTACACCGTGAAGATTCAGAC. 132
386 erglnlyrllleaspaSnalyservalasnservalaglyleuHlthryal 402
133 .....ATGATGAAGCCACTTACTCCAGCAGTCAGATCAT..... 168
403 GlnAlaAlaProAlaGlyleuasnHlProGlnGlnlleasSerHlSe 419
169 .....CAAAACCAAAACAG...GAATCTTCTC 194
419 fargleuProHlSerAlaHlProSerGlylySProProSerAlaProA 436
195 TTAGGAATACATCCACACAGATGAATTTAAACTTTTCTTCATCAGC 244
436 lasevalaProGlnasnValPheSerThrThrValSerSerGlytyrAsn 452
245 CTCCTAAGCTCGAGATATCTTT.....GAGGTGAACGTGCTAAAAATGATTA 293
268 .....GAGGTGAACGTGCTAAAAATGATTA 293
453 ThrlylyslleGlylySargleuasnllleGlnleuylslysglythgl 469
268 .....GAGGTGAACGTGCTAAAAATGATTA 293
469 uGlyleuGlyPheSerllethr.....Sera 478
294 CAGCTTGGGATTAAGTGTACAGTCTGTTGACAAAGGAGTGTGTAA 343
478 rgaSpvalThlileGlylySeralalProIlelyrVallylSasnllleu 494
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

344 CGAGTGTACAGATGTGGC.....ATTATGTGAAGCTGTATT 384
495 ProArgGlyAlaAlaAlaGlnaspGlyArgleuLySargle 511
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
385 CCCAGGAGCAGCAGAGTCTGATGTAGAAATTCACAAAGTCATCGCT 434
511 uileGlyValasnGlyAlaSpLeuValGlySerGlnGlnlyVal 528
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435 CTTGCTGTCAATGAGGTGTCTAGAGAGCACCATAAGCAAGC 484
528 alSerleuLeuArgSerThrLySmetGlnGlyThrValSerleuVal 544
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
485 TGGAAACACTGAGAAATAC.....GGACAGGTGTCTATCTGT 525
545 PheArgGlnGlnaspAlaPheHlSerProArgly..... 555
526 TTGAAAAGGACAAATCTCAACATCTAAAGAACATGTCCGGTACCCC 575
556 .....LeuasnAlaGluProSerGlnMetGlnllePro..... 566
576 ACAGTGTACCTTTCAGATCAGAAATGCCAAGTCAAGGCCAGAAAAG 625
567 ....LySgluThrlySalaGlyaspGlnaspIleValleuThrProasp 581
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
626 TGAAGAAAACACTCAGGTCAAGACTACAGCTTGTG..... 663
582 GlyThrArgGluPheleuThrPheGlyValProleu...AsnaspSerG 597
664 .....ACTGAAGAAATACATTTGAGGTAAATTTATTTAAAAATGAC 707
597 ySerAlaGlyleuGlyValSerValLySlyasnArgSerLySglnasnH 614
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
708 AGCTTACGATTCAGTTTCTCTCGAGAGATTAATCTTATACCGGACAA 757
614 lSAlaaspLeuGlyllePheValySerlleleasnGlyGlyAlaAla 630
758 TTAATGCCACATGTAGGTAAAGGTTCTTCTGACAGCAGCA 807
631 SerlyaspGlyArgleuAlaValasnaspGlnleuAlaAlaValasnG 647
808 GCAGAAAGTGAATAATGTAGTGAAGAGATGTATCTTCAAGGTAAAG 857
647 yGlySerleuLeuGlylySThrAsnGlnaspAlaMetGlnlyrleuAla 664
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
858 AGCCTTTTGAAGGACATATCTCAGCAGAAATCATCTCTCTCAGGG 907
664 rSerMetSerThrGlnGlyasnLySarglyMetlleGlnleuAla 680
908 GAAGTCTCCAGAA.....GTAATCTTGTCTCTC 936
681 AlaArgArglleSerLyScysasnGlyleuLySserProGlySerProP 697
937 TGCAG.....CCTCCACC 950
697 oGlyProGlyleuProIleGlnlyrAlaLeuaspArglylualArgly 714
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
951 TGGTGTCTACCGGAATGTGATCTGCGCT..... 981
714 leSerHlSerleuLySserGlylyleGlnlyleuaspGlySerProSer 730
982 ..TTGACCCCACTTCAGTCTCCAGCACAAATACTTCCAAACACAGTAA 1029
731 ArgasnAlaAlaLeuSerArglle..... 738
1030 GACTTCTTCAGCCATCATGTGTGAGCAACACACAGCTCAGATGAAA 1079
739 ....MetGlylySerGlylySlyrGlnleu...SerProThrValasnM 753
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1080 TGAATGTCTAGACAAAAGCAAAAGCAAGTCCCATCTCAGCAAGAG 1129
753 etProGlnaspArgThrVallleleGlyaspArglyleuProvalleu 769
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1130 ACAATTACGTGACAGCACTGGAGTGTGAGAGATGATCTGTGTACAGCT 1179

```

```

770 ProProHISLeuSerAspGlnSer...SerSerSerSerHisAspAspVa 785
    ||| .....|||.....|||.....|||.....|||.....|||.....
1180 CCGAGAAACATATCAATTCGACCTGACCTTTCATCAGACTCT 1229
785 IGLyPheValThrAlaAspAlaGLyThrTrpAlaLysAlaAlaIleSer. 801
    .....|||.....|||.....|||.....|||.....|||.....
1230 AAGCAACATGTATCAGCAGCAGATCATCATGAAACCCCAAGAGTC 1279
801 .....|||.....|||.....|||.....|||.....|||.....
1280 AAGAATATCATTTTACCATTGTTTACTATCCCTAGAAATTCCTCAAT 1329
802 .....|||.....|||.....|||.....|||.....|||.....
1330 AAGCAGAGCTTTCAGCAGCAGATTCCTCCCTACACACCGATATGCG 1379
813 ProValIleAlaPheGlnArgGluGlyPheGlyArgGlnSerMetSer 829
    ||| .....|||.....|||.....|||.....|||.....
1380 TCCTGGCGAGATTATCAACCCCAATCAGAAATCTGCTTCTCTAGTTG 1428

seq_name: /cgn2_6/ptodata/2/1na/6a_COMB.seq.us-08-931-999-4

seq_documentation_block:
: Sequence 4, Application US/08931999
: Patent No. 6043219
: GENERAL INFORMATION:
: APPLICANT: Iandolo, John J.
: APPLICANT: Crupper, Scott S.
: TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESS: Hovey, Williams, Timmons & Collins
: STREET: 2405 Grand Boulevard, Suite 400
: CITY: Kansas City
: STATE: Missouri
: COUNTRY: U.S.A.
: ZIP: 64108
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/931,999
: FILING DATE:
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/710,561
: FILING DATE: 19-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Collins, John M.
: REGISTRATION NUMBER: 26,262
: REFERENCE/DOCKET NUMBER: 25043-A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 816/474-9050
: TELEFAX: 816/474-9057
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6755 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Staphylococcus aureus
: STRAIN: UT0007
: US-08-931-999-4

```

```

Quality: 244.50      Length: 1306
Ratio: 0.414      Gaps: 56
Percent Similarity: 45.253      Percent Identity: 19.219

Alignment block:
US-09-757-781-2 x US-08-931-999-4 ..

Align seg 1/1 to: US-08-931-999-4 from: 1 to: 6755

154 SerValSerAspSerAsnPheSerSerGluGluProSerArgLysAsnPr 170
    ||| .....|||.....|||.....|||.....|||.....
610 AGCAGAGCAATACCAAGAAAGAGGGGAGGCACCAACGAAAAAAG 659
170 otherArgPheSerThrAlaGlyPheLeuLysGlnAsnThrAla.... 185
    |||.....|||.....|||.....|||.....|||.....
660 AACAAAA...AAGAAAGAACCGGAAAGCGCAAGAAAAAAGCAAGCAAA 706
186 .....|||.....|||.....|||.....|||.....
707 AAGCGCAAAAGCGCCACCAAGAAAAACACCCCAAAAGGAGACAGA 756
194 LysLysAspGluAsnTyrArgSerLeuProArgAspThrSerAsnTrpSe 210
    |||.....|||.....|||.....|||.....|||.....
757 AAAAAAAGCAAGCAAAAGAAAAA.....CGAAGCAACAGAAACCCAG 800
210 raen.....GlnPheGlnArgAspAsnAlaArgSerSerSerLys 225
    |||.....|||.....|||.....|||.....|||.....
801 CAACAAAAAAGACGCAAGAAAGACGCAAGAAAAAACAACACACCCCA 850
225 erHisProMetValGlyLysTrpLeuGluLysGlnGluGlnAspGluAsp 241
    ..|||.....|||.....|||.....|||.....
851 AAACAAAGAAAGCAACCAAA.....AAAAAGCAAAACCAACCAAAAG 894
242 GlyThrGluGluAspAsnSerArgValGluProValGlyHisAlaAspTh 258
    |||.....|||.....|||.....|||.....|||.....
895 GGCACGCAAGAGAGAGAAAAAGC..... 918
258 rGlyLeuGluHisIleProAsnPheSerLeuAspMetValLysLeuV 275
    |||.....|||.....|||.....|||.....|||.....
919 GGAAGGAAAAAAGCAACCA..... 936
275 alGluValProAsnAspGlyGlyProLeuGlyIleHisValValProPhe 291
936 .....|||.....|||.....|||.....|||.....
292 SerAlaArgGlyGlyArgThrLeuGlyLeuValLysArgLeuLul 308
    |||.....|||.....|||.....|||.....|||.....
937 .....AGGAAGGACCAAAAGAAAGAAAAAAGAAAAAAGCGAAGAA 980
308 ysgLysGlyLysAlaGluHisGluAsnLeuPheArgGluAsnAspCysIle 324
    |||.....|||.....|||.....|||.....|||.....
981 CAGGAAAAAAGAAAGCAAAACAGAGAAAGCAAAACGAAACACCG 1030
325 ValArgIle.....AsnAspGlyAspLeuArgAsnAlaGlyPhe 337
    .....|||.....|||.....|||.....|||.....
1031 ACCGAAACCAAAAAAAGAAAGCAAGCCG.....AGGAAAGAGA 1074
337 eGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThrProIleIle 354
    .....|||.....|||.....|||.....|||.....
1075 CAGAAAAAAGCAAGAGACACAGCGAAGAGAAAC..... 1114
354 rPheHisValValProAlaAla..... 361
    |||.....|||.....|||.....|||.....
1115 .....CACAAAGAAACAAAGCAAGAAAGAAAAAAGCAAGCAACAA 1159
362 .....AsnLysGluGlnTyrGluGlnLeuSerGlnSerGly 374
    .....|||.....|||.....|||.....|||.....
1160 CCCCAGAAAGAAAGAAAGAAAGCAAGAGAAAGCAAGCAAGAAAGAA 1209
374 sAsnAsnTyrTyrSerSerArg..... 381
    .....|||.....|||.....|||.....|||.....
1210 AAAAAAGCGGCAAAAGCAAGACCAAGACAAAGCAAGAAAGAAAGAA 1259

```

alignment\_scores:

Thu Jul 25 08:38:38 2002

```
382 ..... pheSerProAspSerGlnTyrIleAsp 390
1260 GAGGACGAAAGAAAAAACAGACACCAAAAGAACGAAAGAAA 1309
391 AsnArgSerValAsnSerAlaGlyLeuHisThrValGlnArg .....Al 405
.....
1310 AGAAGACCGGAAACCGCAAAACGAGAGACCGCCAAAGAGAGAGAG 1359
405 AspArgLeu.....AsnHisProProGlnGlnIleAspSer. 417
.....
1360 ACCCAGAGAGAGCGAAGAAAGGCGAAGCGCCACCAACGAAACCA 1409
418 .....HisSerArgLeuPro 422
.....
1410 AAAAAAACCAAAAGAAAGAGCGAGGAAAGCAAAAGCAAAAGAGAAA 1459
423 HisSerAlaHisProSerGlyLysProProSerAlaProAlaSerAlaPr 439
.....
1460 CAACGACACAGAAACACCGCGAAGAGACCGCAAGCGCAAGAGCC.. 1507
439 GlnAsnValPheSerThrThrValSerSerGlyTyrAsnThrLysLysI 456
.....
1508 .....AACGCGAAAGACACAGAAACCAAGAAAGAAA 1538
456 IeGlyLysArgLeuAsnIleGlnLeuLysLysGlyThrGlnGlyLeuGly 472
.....
1539 ACAGCAAAAGCGAAAGAGAAAGAAAGAAAG..... 1570
473 PheSerIleThrSerArgAspValThrIleGlyLysAlaProIleTyr 489
.....
1571 .....AAGAAAGAGCGGACGAGAGACGAGGAGCGGAGCAAAA 1608
489 rValLysAsnIleLeuProArgGlyAlaAlaIleGlnAspArgLeu 506
.....
1609 GAGAAAGAAC.....CAAGAAAGACCGACGA 1634
506 ySAlaGlyAspArgLeuIleGlnValAsnGlyLysAspLeuValGlyLys 522
.....
1635 AATCAAAAGAAAGAA.....GGAAG 1654
523 SerGlnGluGlnValValSerLeuLeuArgSerThrLysMetGlnGlyTh 539
.....
1655 AAGCAAAAG.....AAAAAGAAAGAACACCGCGAAG 1686
539 rValSerLeuLeuValPheArgGlnGluAspAlaPheHisProArgGlu 556
.....
1687 CCAGGAAAGCCGAGAGAGCAAAAGAAAGCAAAAGCAAAAGAGAA 1736
556 euAsnAlaGluProSerGlnMetGlnIleProLysGluThrLysAlaGlu 572
.....
1737 AATAGAGAACAAAGAAACAAAGCAAGCAAGCAAAAGAAAGCGGAA 1786
573 AspGluAspIleValLeuThrProAspGlyThrArgGluPheLeuThrPh 589
.....
1787 GACAAA..... 1792
589 eGluValProLeuAsnAspSerGlySerAlaGlyLeuGlyValSerValL 606
.....
1793 .....A 1793
606 ySGLysAsnArgSerLysGluAsnHisAlaAspLeuGlyIlePheValLys 622
.....
1794 AATACACGCGGCAAGACCGGACGACGCAAGCAAGAG..... 1832
622 sSerIleLeuAsnGlyAlaAlaSerLysAspGlyArgLeuArgValA 639
.....
1833 .....AAAGCGGAACAAAGAGAAAGAAAGAAAGCAAAAGAA 1872
639 snAspGlnLeuIleAlaValAsnGlyGluSerLeuLeuGlyLysThrAsn 655
.....
1873 ACCAGAGCGGACCGGACGAGGAGGAGAGCAAAAGCAAAAGCGGAA 1922
656 GlnAspAlaMetGluThrLeuArgArgSerMetSerThrGluGlyAsnLys 672
```

us-09-757-781-2.rn1

```
1923 AAG...AAACACGAAAGAAAGAAAGAAAGCAAGCGCGCAAGAAAGAAACA 1969
672 sArgGlyMetIleGlnLeuIleValAlaArgArgIleSerLysCysAsnG 689
.....
1970 AAGAGGA.....GAGG 1980
689 IuLeuLysSerProGlySerProProGlyProGluLeuProIleGluThr 705
.....
1981 AGAATACACGCGCGAG..... 1997
706 AlaLeuAspAspArgGluArgArgIleSerHisSerLeuTyrSerGlyI 722
.....
1998 .....AAGACAGCAAAACCGGA..... 2015
722 eGlnGlyLeuAspGluSerProSerArgAsnAlaAlaLeuSerArgIle 739
.....
2016 .....GAACAGAAAGACAGAGAAAGAGCGCCCAACAGAGGA 2055
739 eGlyGluSerGlyLysTyrGlnLeuSerProThrValAsnMetProGln 755
.....
2056 AAGCAAAAGAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAG 2078
756 AspAspThrValIleIleGluAspAspArgLeuProValLeuProProH 772
.....
2079 .....SerSerSerSerHisAspA 784
772 sLeuSerAspGlnSer.....SerSerSerSerHisAspA 784
.....
2090 CAAGCGCAAAACAGAAAGCGAGCGGAAAGCAAGCAAGCAAGCGCAC 2139
784 sPValGlyPheValThrAlaAspAlaGlyThrThrPAlaLysAlaIle 800
.....
2140 ACAGAAAGAAAGAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2189
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeu 817
.....
2190 GAACCAAAAGAGAAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAG 2231
817 aPheGlnArgGlnGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
.....
2232 .....AAGCGCGAAAGCGGACGAGAGAGCA 2259
834 ySGLnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
.....
2260 AGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 859
851 LysSerMetAspLeuGlyIleAlaAsp..... 859
.....
2310 AGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 873
860 .....GluThrLysLeuAsnThrValAspAspGlnLysAlaGlyS 873
.....
2348 CGGACAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2397
873 eProSerArgAspValGlyProSerLeuGlyLeuLys.....Lys 886
.....
2398 GGAAGCAAAACACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2447
887 SerSerSerLeuGluSerLeuGlnThrAlaValAlaGluValThrLeuAs 903
.....
2448 CCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2486
903 nGlyAspIleProPheHisArgProArgProArgIleIleArgGlyArg 920
.....
2487 .....AACGCAACAAAGAAAGAAAGAGAAAGAAAGAGAA 2514
920 LysCysAsnGluSerPheArgAlaAlaIleAspLysSerTyrAspLysPro 936
.....
2515 GGGCAAGCAAGAG.....CGCAAAAGACCGCGCGCA 2549
937 AlaValAspAspAspAspGlnGlyMetGluThrLeuGlnGluAspThrG 953
.....
```

```

2550 GGA.....GAAACGAGAGAAAGCAACCGCCAGCAAGAAAGGCCCA 2590
953 uGUSERSerArgSerGly.....ArgGUSERValSerT 965
      ::::: ||| ||| :::::
2591 CAGGACGACAGAGAGAGAGAGAGAAACCAAGCAAGCAAGAAAGAGAA 2640
965 hVALSerAspGlnProSerHisSerLeuGlnArgGlnMetLsnGlyAsn 981
      ||||| ::::: ||||| :::::
2641 CAGCAGAGCAAGAAACCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGCC 2690
982 GlnGlnGlyAspLysThrAspArgGlyLys.....AspLys 994
      ::::: ||| ||||| ||||| |||||
2691 AAGAGACCA...CAGCAGACAAAGAAAGAGAGAGAGAGAGAGAGAGAG 2737
994 sThrGly...LysGlnLysLysLysAspArgAspLysGlnLysAspLysM 1010
      | ||| ::::: ||| ||||| ||||| |||||
2738 GAGAGGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2787
1010 eLysAlaLysLysGly.....MetLeuLysGlyLeuGlyAsp 1022
      ||||| ::::: ||| ::::: |||||
2788 GGGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2837
1023 MetPheArgPheGlyLysHisArgLysAspAspLysIleGlnLysThrG 1039
2838 .....AGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2863
1039 LysIleLysIleGln...GlnSerPheThrSerGlnGlnGlnGlnGln 1055
      |::: ||| ||| ::::: ||||| :::::
2864 CAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2913
1055 rGMetLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1071
      :: ||| ||||| ::::: ||| ::::: |||||
2914 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2960
1072 GluArgGlnAlaArgGlnArgGlnArgGlnGlnGlnGlnGlnGlnGln 1088
      ||||| ::::: ||||| ||||| |||||
2961 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3007
1088 gThrPheGlyCysAspAspGlnLeuMetLysGlyLysValSerSerTyrG 1105
3008 AAAAAAGAGG.....GACACAGAGCA 3027
1105 LUGLysMetAla..LeuAsnAlaArgProGlnSerProArgGlnGlnH 1121
      ::||| ||| ::::: ||||| ||||| |||||
3028 GGGGAGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3077
1121 sMetMetAspAlaLeuTyrAlaGlnValLysLysProArgAsnSerLys 1138
      ::::: ||||| ::::: ||||| |||||
3078 GCAAGAGGGGGAG.....AGAAAGCAAGAAAGAGAGAGAGAGAGAG 3112
1138 rOserProValAspSerAsnArgSerThrProSerAsnHisAspArg... 1153
      ::::: ||||| ||||| ||||| |||||
3113 GACACAGAGCA...ACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3159
1154 .....IleGlnArgLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGln 1168
      ||||| ::::: ||||| ||||| |||||
3160 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3209
1168 uAspValGlnAspArgArgArgThrTyrSerPheGlnGlnGlnGlnGln 1185
      ::||| ||| ::::: ||||| |||||
3210 GCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3259
1185 sn...AlaArgProAlaThrGln...SerGlyArgHisSerValSerVal 1199
      || ||||| ||||| ||||| :::::
3260 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
1200 GluValGlnMetGlnArgGlnArgGlnGlnGlnGlnGlnGlnGlnGln 1216
      ::::: ||||| ||||| ||||| |||||
3301 .....CGAAGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3344
1216 nAlaGlnArgGlnTyrSerSerLeuProArgGlnSerArgLysAsnAla 1233
      : ::::: ||||| ||||| ||||| |||||
3345 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3394

```

```

1233 eSerValSerGlnAspSerTyrGlnGlnAsnTyrSer...ProGlyGln 1248
      ::::: ||||| ::::: ||||| |||||
3395 GCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3444
1249 GlyPheGlnSerAlaLysGlnAsnProArgTyrSerSerTyrGlnGly 1265
      ||| ::::: ||||| ||||| :::::
3445 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3494
1265 rArgAsnGlyTyrLeuGlyLysHisGlyPheAsnAlaArgValMetLeu 1282
3495 AAGGGH.....AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3520
1282 LurThrGlnLeuLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1298
      ::||| ::::: ||||| ||||| ||||| |||||
3521 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3570
1299 LysLysGlnProProSerGlnGlyProSerAsnTyrAspSerTyrLys 1315
      ::::: ||||| ::::: ||||| |||||
3571 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3620
1315 sValGlnAspProSerTyrAlaProProLysGlyProPheArgGlnAsp 1332
3621 ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3670
1332 aLProProSerProSerGlnValAlaArgLeuAsnArgLeuGlnPro 1348
3671 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
1349 GlnLysGlyArg 1352
3721 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3732
seq_name: /cgn2_6/ptodata/2/lna/6A_COMP.seq:us-09-045-632-1
seq_documentation_block:
; Sequence 1, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: Hugenltz, Richard L.
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4537 base pairs

```

```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-045-632-1

```

```

alignment_scores:
  Quality: 233.50      Length: 1056
  Ratio: 0.509        Gaps: 43
  Percent Similarity: 43.466   Percent Identity: 19.697
alignment_block:
US-09-757-781-2 x US-09-045-632-1

```

Align seg 1/1 to: US-09-045-632-1 from: 1 to: 4527

```

215 ArgAspAsnAlaArgSerSerLeuSerAlaSerHisProMetValGlyLys 231
|||
685 CGAGGGGGAGCCGATGATGACAGAAATAAATCCCGCTGCGATGATGATAC 734
231 sTrpLeuGluLys...GlnGluGlnAspGlyLysProGluGluAsp 247
|||||
735 CTGTGTCCTGCTGGAGGGCCCTGCTGACAGAGAGGAGCCATCAACCTG 784
247 snSerArgValGluProVal...GlyHisAlaSprThrGlyLeuGluHis 262
|||||
785 GAGACAGGTGCTCAGTGTGATGAAATTCGGCTTCGGGAACCAACCCAT 834
263 IleProAsnPheSerLeu.....AspAspMetValLysLe 274
|||||
835 GCTGAGGCCATGAGCATCTTAAACAGTGGAGCAAGAACCAACCTGCT 884
274 ValGlu.....ValProAsnAspGlyLys 283
|||||
885 GATGATATGATGATGCTCCGTGATGATGATCTGTGACCAACAGATCCGGCG 934
283 rGluGlyIleHisValValProPheSerAlaArgGlyLysArgThrLeu 299
|||||
935 CACTA.....CTAGTTGAAGTTGCCAAACCTCCGGGTGCAACGCTT 975
300 Gly.....LeuLeuVal 303
976 GGGGTTCACATACTACCTCCGTGCTGATMAACAAACAGTCAATCTCAT 1025
303 LysArgLeuGluLysGlyLysAlaGluHisGluAsnLeuPheArg 320
|||||
1026 AGACAAATCAAAATCTGCACATTCGGGACAGGTGCGGGGGCTTACATG 1075
320 LuAsnAspCysIleValArgIleAsnAspGlyAspLeuArgAsnArg 336
|||||
1076 TGGGAGACCAATCTCTCCATCGACGCAACGATGAGTACTGTACC 1125
337 PheGluGlnAla.....ProSerGlnMetGln 340
|||||
1126 CTCGCAGAACCAACCACTTCTGCGCAACACACTGACAGGCAAGCT 1175
341 .....GlnHisMetPheArgGlnAlaMetArgThrProIle 1223
1176 GGGAGATTCTCCACACATCAAGACCCGCTGCGCTGAGGAGCACTGA.. 1223
353 LeTrpPheHisValValProAlaAlaAsnLysGluGlnArgGlnLeu 369
|||||
1224 .....CATGTCAAAAATTCAGAGGAGCGACAG..... 1250
370 SerGlnSerGluLysAsnAsnTrpTyrSerSerArgPheSerProAspSe 386
1250 .....GluAsp.....GluAspIleVal 577
386 rGlnTrpTyrLeuAspAsnArgSerValAsnSerAlaGlyLeuHisThrValG 403
1251 .....ACAGCTTCCTGGGATCCCTGGGCGACAGCAAGCAAGTCC 1287

```

```

403 LnaArgAla..ProArgLeuAsnHisProProGluGlnIleAspSerHisSe 419
|||||
1288 AGCGTTCATATACCAACCATACCATATACCCGACCCACGAC...CATGG 1334
419 rArgLeuProHisSerAlaHisProSerGlyLysProProSerAlaPro.. 435
|||||
1335 CCGAGTACACAGCCCTGGGTTCGGAAGCGCTTACTCCAAACAGCCCTC 1384
436 .....AlaSerAlaProGlnAsnValPheSerThrVal 447
1385 CGCGATGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1434
448 Ser.....SerGlyTyrAsnThrLysLys 455
|||||
1435 AGTTCCTGGAACATGGGAGCTTACTCTGACGACCTCTACTCTACTGACC 1484
455 sIleGlyLysArgLeuAsnIleGlnLeuLysLysGlyThrGluGluLeuG 472
|||||
1485 ACAGAGAACCATGATGAGAGAGAGACTGAAAAAGAGACTTCAAAAGCT 1534
472 LysPheSerIleThrSerArgAspValThrIleGlyLys..... 484
1535 CACTGCTTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1584
484 .....ACTGAACACACAGAGTTGCTGACGCTGATCTGACCTGCTTGG 1634
485 .....SerAlaProI 488
1635 AATCCACCTGAGGAGCGCTGTTGGCAACAGACTCTCTCTCCAC 1684
488 LeTyrValLysAsnIleLeuProArgGlyAlaIleGlnAspGlyArg 504
|||||
1685 CTGTGATTCCTATTTGAAGCTGACAGGCCGACAGAGAGATGGTGTG 1734
505 LeuLysAlaGluLysPheLeuIleGluValAsnGlyValAspLeuValG 521
|||||
1735 CTGCACATTTGAGACAGACAGTCAATGATTAATGAAATCCCAACAGAGA 1784
521 LysSerGlnGluGluValAlaSerLeuLeuArgSerThrLysMetGlu 538
|||||
1785 CAGCACCTTTGAGAGCAATCAAGCTCTCTCTCTCTCTCTCTCTCTCT 1834
538 LysThrValSerLeu.....LeuValPheArg 546
|||||
1835 GCAAGTCACTGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1884
547 GlnGluAspAlaPheHis.....ProArgGluLeuAsnAlaGlu.. 559
|||||
1885 AGTATGGAACATTTCAATGTAACCTGCTTAAGAACACAGAGCTGAGCT 1934
560 .....ProSerGlnMetGln..... 564
1935 TGGAAATTAACCATCAGTTCGCAATCCAGTAAGAAACACAGGAGACCTCTCG 1984
565 .....IleProLysGluThrLysAla..... 571
1985 TCATTTCAATATCAAGAAAGGACAGTGTGGACACACAGAACTGGAACCTCG 2034
571 .....GluAsp.....GluAspIleVal 577
2035 GAACCTTGAGATAAATTCCTGCGCATGATACATCCGTTGGACAGCTG 2084
572 .....GluAsp.....GluAspIleVal 577
2085 TTCCATGGAAGATGACATCCAGATCTCCAGAGTGTGAAGACTGTGGA 2134
578 .....LeuThrProAspGlyThrArgGlu..... 585
2135 AGCTCAAAATCCCAAAAGATGAAGATTAATCTCAGATGAGCAAGAGATTC 2184
586 ...PheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGlyLe 601

```



```

2185 GGGCGCATTTTATACAGCGGTGAGAGCCCTGATGGGGGCC...CT 2231
2186 ugiValSerValIysgluValasnArgSerIysgluAsnHisAlaAspLeu 618
2232 TGGCATCACCATTTCGTGAAACG.....GAAGAGCCCTTGAT...C 2259
618 LyllePheValIysSerIleIleasnIylAlaIalaSerIysAspGly 634
2270 CTAATTATCATCTCGAGCCCTCAATAAGGGGATAGCTGAAGAGCTGGA 2319
635 ArgLeuArgValAsnAspGlnLeuIleAlaValAsnIylSerIleule 651
2320 GGCATCCACATCGAGATCGAATCCATCTAGCCATCATATGACAGTCTGAA 2369
651 ugiLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSer 668
2370 GGGGAAGCCCTCTGAGTGAAGCCATCCACTGCTACAGATGGGGGAGGA 2419
668 hrgLysLysnLysArgGlyMetIleGlnLeuIleValAlaArgArgIle 684
2420 CT.....GTACCCCTGAAATTT 2436
685 SerLysCysAsnGlnLeuLysSerProGlySerProGlyProGlyule 701
2437 AAGAGCAGACAGATGCTCAACCTGCCCTCAAGTCC.....AAGAAGCT 2480
701 uProlIleGluThrAlaLeuAspArgGluArgArgIleSerHisSerI 718
2481 GCCCATCCCC.....AGCCACTCAA 2500
718 eufIysSerGlyIleGluGlyLeuAspGlnSerPro...SerArgAsnAla 733
2501 GTGACCTGATGATGGTAGAGAGAGCCCTCCCAATACAAAGGCTGCG 2550
734 AlaLeuSerArgIle..... 738
2551 AAGCTCTGTACGTGATACCCCTCCACAGTACCCAGCGTGAGAGTGTGT 2600
739 .....MetGlyLysSerGlyLysTYrGlnLeuSerProT 750
2601 GGACTCCTGGGAGCGATCTGGAATAGTCCAGGTATGGAGTCAAGCA 2650
750 hValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeu 766
2651 CGACTTTCAGACTCAGGGTACATTTCAACACCTATGACTGAGAGAGT 2700
767 ProValLeuProProHisLeuSer.....AspGlnSerSerSerSer 781
2701 CCGAAGAAAAGACCAAGCCGTGCCCCGCGTCCCAAGCCTCGAAGCCAGAC 2750
781 rHisAspAspValAlaGlyPheValThrAlaAspAlaGlyThrTrpAlaLysA 798
2751 ATACCCAGATGTGGCTTGAGTATATGAAGAC.....TGGGACCGGT 2791
798 IalaIleSerAspSerAlaAspCysSerLeuSerProAspValAspPro 814
2792 CCACAGCGCACTGCTTCGCGGGGCTTCTGACAGTGCAGACGCTGAA... 2838
815 ValLeuAlaIleGlnArgGlyLysPhe..... 823
2839 .....CAAGAGAAACCTTGTGTCACACCATTTGAGAGACCT 2876
823 ..... 823
2877 GGAGACCTCGGGCGCAAGTCCGGAATCTGAGAGAGCTCGAGCAACAATCA 2926
823 ..... 823
2927 TGTGCGGAGCACTATGATTGATCATGAGGCTCAACGCTCGCAGT 2976
824 .....GlyArgGlnSerMetSerGlnLysArgThrLysGlnPheSerAs 838
|||||

```

```

2977 CAGCTGGGGCGCAGAGCCAGCTTCCAGAACGAGCAAT..... 3015
838 PALSerGlnLeuAspPheValIysThrArgLysSerIysSerMet... 853
3016 ....TTCAGGCGCACATATACCAAAAGCATCTGCGACAGCAACTGCCCC 3061
854 ..AspLeuGlyIleAlaAspGlnThrLysLeuAsnThrValAspAsp... 868
3062 CGGATGTGGCGCAAGAGCTGTGACCTCGGAAATAATGACAAAGAAATC 3111
869 GlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyLeuLys 885
3112 AAGGAGATTATGTCCCAACCTCCGCTGAGACCTTCAAGAGTGCATTATA 3161
885 sLysSerSerSerLeuGlnSerIleGlnThrAlaValAlaGluValThrL 902
3162 CAAGACTCTGCGCATGAGAGACTTGGGTTCAGTGTGCAATGCGCCGTC 3211
902 euAsn.....GlyAsp 905
3212 TGAGAAAAGGGGTATGTCAAAATATCCGCCAGCTGGGCCAGGTGAT 3261
906 IleProPheHisArgProArgProArgIleIleArgGlyArgGlyCysAs 922
3262 CTTCGAGGCTTGAAAGCCTTATGACAGAGCTCTTACAG.....GTTAA 3302
922 nGlnSerPheArgAlaIalaIleAspLysSerIysAspLysProAlaValA 939
3303 CCAGCTCGAAGCAGAGACTTCGACTGCTGCTGTGCTGCTGCTCATAG 3352
939 sPAspAspAspGlnGlyMetGlnThrLeuGlnGluAspThrGlnGlnSer 955
3353 CCGAATCCGGTAAACAGCTGAGCTGTTATTATACAGAAACCCCTGACC 3402
956 SerArgSerGlyArgGlnSerValSerThrAlaSerAsp..... 968
3403 TCCCAAGATGATAGAACACCGCGCTCTGCCAGTGCAGACGAGACA 3452
969 .....GlnProSerHisSerLeuGlnArgGlnMetA 979
3453 GAACAGGCTTCTTCAGCAGACCCAGCCAC.....G 3484
979 snGlyAsnGlnGlnLysGlyAspLysThrAsp..... 989
3485 GTGTAACCTAGAGACAGACAGAACCCACTAATACATTAATTAATTT 3534
990 .....ArgLysLysAspLysThrGlyLysGlnLysLysLysA 1002
3535 TATAAGCAGAGCAGAAAGCATATCTACATGATGCTTAAACAAAGAGA 3584
1002 sPArgAspLysGlu 1006
3585 ACAGAACAGAACAG 3598

seq_name: /cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-056-200-93
seq_documentation_block:
; Sequence 93, Application us/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; NUMBER OF INVENTIONS: 117
; METHODS OF USING SAME
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA

```

alignment\_scores:

alignment\_block:

304 LVSARAI ENCLOSURE, 1 to: 9551

[illegible]

337 hegluglnalaglnhisMot ph-  
.....cagcIAGcIAGGAGGAGGAGC 4498

|||||::: ||||  
4499 AGGAGACCCCCC ::||| ||| 353

.....  
..GAGCCGATTAGAGCCCATCCGAGTGGCAGTGC /518

353 eTrpPheHisValAlaProAlaAlaAsnLys...[Glu]nTurct...

[illegible]

369 euSergIInSergIuIwbaa...  
...GCAAGICTTACTCGAGGC 4598

::::| | TyrSerSerArgPheSerProasp 385  
 4599 CCCCCC

[illegible]

386 SerGlnTyrIle.....

4649 GGGCCGAGGTGACCTGCATATGGCAGCAAGCGA...AS 390

...ccnccnccgaggaacccggctcaccggcag 4698

Page 22

5361 ..... ||| |||:|||||: |||  
5385 GAAAAAATACG.....CGAGAGAGAGAGCG.....GGA 5384  
699 ProGluLeuProIleGluThrAlaLeuAspArgGluArgIleSe 715  
715 rHisSerLeuTySerGlyIleGlu.....GlyLeuAspGluSerP 729  
5414 ACAGAGAGAGAGACGCTGCTGAGAGAGAACGGAGAAAGAGCCG 5463  
729 ro.....SerArg 731  
5464 CAGAGAGTGGAGAGCAGTACCGCAAAAAAGACAGCTCAGCAGAGAG 5513  
732 AAAAAA ||| |||:|||||: |||  
5514 AGAGCAGCTGCTGAGAGAGAACGGAGAAAG..... 5546  
748 rProThrValAsnMetProGlnAspArgThrValIle..... 760  
5547 .....AAGACTCCAGAGCGGAGAGCGCAATATCGGAGAGAG 5586  
761 .....IleGluAspArg 764  
5587 GAGCTGCACAGAGAGAGAGCAGCTGCTGGAGAGAACGGAGAGCAG 5636  
765 ArgLeuProValLeuProProHisSerAspGlnSerSerSerSe 781  
5637 AAGCGCCAGAGAGCTGAGAGAGCAATATCGAAAGAGAGAGAGCTCAGC 5686  
781 rHisAspArgValGlyPheValThrAlaAspAlaGlyThrTrpAlaLys 798  
5687 AGCA.....GGAAAGACAGCTCTGAGAGAGAGAACCGGAGAGAG 5727  
798 LaLaIleSerAspSerAlaAspCys.....SerLeuSerP 810  
5728 AGCGCCGAGAGAGAGAGAGCAATGTCGGAGAGAGAGAGAGCTGAGCA 5777  
810 oAspValAspProValLeuAlaPheGlnArgGluGlyPheGlyArgGln 827  
5778 GGAGAGAGAGAGAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 5827  
827 erMetSer.....GluLysArg 832  
5828 ACTGAGAGAGCAATATCGGAGAGAGAGAGAGCTTACGCCAGAAAG 5877  
833 ThrLysGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArg 849  
5878 AAGACAGCATACCGGAGATGATCAGCGCAGATGATGAAATGGCAG.. 5925  
849 sSerLysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThr 866  
5926 .....TGGGAACCAAAAAAAGAAATTCAG 5950  
866 aAspAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeu 882  
5951 TTCGTGATGAACAAGTTTACTGCAAGAGCAAGAGATGAACAGTTCCG 6000  
883 GlyLeuLysLysSer.....SerSerLeuGlnSerLeuGln 894  
894 nThrAlaValAlaGluValThrLeuAsnGlyAspIleProPheHisArg 911  
6001 CAGTTGGAAAGATTCCAGAGTGGCGAGAGCAATATCCAGCAAGATCTGCA 6050  
6051 GCACCTGCTGGGTGA.....CAGCAAG 6073  
911 roArgProArgIleIleArgGlyArgGlyCysAsnGlnSerPheArg... 926  
6074 ACAGAGATGCTGTGAGCAAGAGAGAGCGCTGCGCAGAGAGCAACAGCAT 6123  
927 .....AlaAlaIleAspLysSerTyAspLysProAlaValAs 939

6124 TTCACAGAGAGAGAACAGCTGAGCGAGAGAAAGCAAAAAAGAACCAAAAG 6173  
939 pAspAsp.....AspGluGlyMetGluThrLeuGluLysPheArg 953  
6174 GCCCGACAGAGACTCCCAAGAGAGAGAAAGCAAGTTGCTGAGAGAGAAAGAG 6223  
953 Lu.....GluSerSerArgSerGlyArgGluSerVal 963  
6224 AAGAGAAAGAGACCGCGTCAGAGACAGCAAGAAATTCGCCGAGAGAGAA 6273  
964 SerThrAlaSerAspGlnProSerHisSerLeuGlnArgGlnMetAsnGln 980  
6274 CAGCTGCTCCAGGAAGGAGGAACAGCCGCTGCTCCGACAGAGAGCTGA 6323  
980 Y.....AsnGlnLysGlyAspLysT 988  
6324 CAGAAATTCGCCGAGAGAGAGAGACTGCTTCATCAGAAACAAAGGAGAAAT 6373  
988 hrAspArgLysLysAspLysThrGlyLysGluLysLysAspArgAsp 1004  
6374 TCCTCGAGAGAGAAACAGCGGCTGCGGAGAGAGAGAGAGAAATTCCTT 6423  
1005 LysGluLysAspLysMetLysAlaLysGlyMetLeuLysGlyLeuGln 1021  
6424 AAGGAGAGACAGCAGCTGCGCTCGAGAGAG.....CG 6455  
1021 yAspMetPheArgPheGlyLysHisArgLys...AspAspLysIleGlu 1037  
6456 CAGACAACTGCTCAGAGACCGCGCACAGAAATTCGCCGAGAGAGAAACAGC 6505  
1037 yThrGlyLysIleLysIleGlnLysSerPheThrSerGluGluLysArg 1053  
6506 AGCTGAGCCGCCAAAGAGCTGACAGAGAAATTCGTGAAGAGAGAGAGAGAG 6555  
1054 IleArg.....MetLysGlnGluGlnLysArg 1062  
6556 GTGCGCCGCGCAGAGAACGAGAGAGAAATTCGTGAGAGAGAGAGAGAGCT 6605  
1062 gIleGlnAlaLysThrArgGluPheArgGlu..... 1072  
6606 GCCCGAGAGAGCTCAGCAAGAAATTCGCCGAGAGAGAGAGAGAGAGAG 6655  
1073 .....ArgGlnAlaArgGluArgAspTyra 1081  
6656 AAGAGGAAGAACAGACAGCTCAGCGCCAGAGAGAGAGAGAGAGAGAGAG 6705  
1082 GlnIleGlnAspPheHisArgThrPheGlyCysAspAspGluLeuMetTy 1098  
6706 GAGGAGAGACACAGCTGCGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAG 6755  
1098 rGlyGlyValSerSerTyArgGluLysSerMetAla...LeuAsnAlaArg 1114  
6756 ACAGAGAACTGCGCAGTACAGAACAGAGAGAGAAATTCCTGAGAGAGAGAG 6805  
1114 roGlnSerProArgGluGlyHisMetLeuAspAlaLeuTyraGlnVal 1130  
6806 AGCAGCTGCACCCGACAGCAACGCGCAGAGAAATTCCTCCAGAGAGAGAG 6855  
1131 LysLysProArgAsnSerLysProSerProValAspSerAsnArgSerThr 1147  
6856 CAGCTGCGCGCG.....CAGAGCGCGGGCA 6881  
1147 rProSerAsnHisAspArgIleGlnArgLeuArgGlnGluPheGln... 1162  
6882 ACAGCGCGCTCAGAGACCGCTGACAGAAATTCGCCGAGAGAGAGAGAGAG 6931  
1163 ..GlnAlaLysGlnAspGluAspValGluAspArgArgThrTySer 1178  
6932 GCCAG 6981  
1179 PheGluGlnProTrpProAsnAlaArgProAlaThrLysSerGlyArgH 1195  
6982 TTCCTTTAGAGAGAGAGAAAGTGCAGCGCCAGAGAGAGAGAG...AGAAA 7028

```

1195 sServalServalGluValGlnmerGlnArgGln..... 1206
      : : : : : ||| ||| : : : : : ||| |||
7029 ATTCATGAGAGAGACGACACAGCTGCGCCGCGAGAGGCGCCACACAGC 7078
      : : : : : ||| ||| : : : : : ||| |||
1207 ..ArgGlnGlnGlnArgGlnSerSerGlnGlnGln..... 1218
      : : : : : ||| ||| : : : : : ||| |||
7079 TCGCGCAGAGAGAGACGACGAAATTCGCCGAGACGACGCTGCTCCAGAA 7128
      : : : : : ||| ||| : : : : : ||| |||
1219 ArgGlnTyrSerSerLeuProArgGlnSerArg..... 1229
      : : : : : ||| ||| : : : : : ||| |||
7129 AGGGAAGAACAGACAGCTGACCGCGCAGAGAGCTGACGAAATTCCTCGA 7178
      : : : : : ||| ||| : : : : : ||| |||
1229 ..... 1229
7179 GGAGGAAACCGAGCTGCGCGCGCAGAGCGCGAACAACAGCTGCTCAGC 7228
      : : : : : ||| ||| : : : : : ||| |||
1230 ..... 1230
7229 ACCCGCAGAGAAATTCCTGAGAGAGAACAGCTGCTCCAGAGAGGCGAG 7278
      : : : : : ||| ||| : : : : : ||| |||
1241 GlnGlnAsnTyrSerProGlnGlnGlnGlnSerAlaLysGlnAsnPr 1257
      : : : : : ||| ||| : : : : : ||| |||
7279 GAACAGCAGCTGCGCGCGCAGAGAGCTGACGAAATTCGCCGAGAGAGA 7328
      : : : : : ||| ||| : : : : : ||| |||
1257 oArgTyrSerSerTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1274
      : : : : : ||| ||| : : : : : ||| |||
7329 ACAGCAGCTCGCGCGCGCAGAGAGAGAGAAATTCCTC..... 7368
      : : : : : ||| ||| : : : : : ||| |||
1274 LysPheAsnAlaArgValMetLeuGlnTyrGlnGlnGlnGlnGlnGln 1290
      : : : : : ||| ||| : : : : : ||| |||
7369 ..... 7369
1291 GlnArgArgLys.....GlnGlnGlnMetLysLysGlnProProse 1304
      : : : : : ||| ||| : : : : : ||| |||
7399 CCGGAGAGAGAAATTCCTGAGAGAGAACAGCTGCGCCAAACGAGCA 7448
      : : : : : ||| ||| : : : : : ||| |||
1304 rGlnLysProSerAsnTyrAspSerTyrLysValGlnAspProSerT 1321
      : : : : : ||| ||| : : : : : ||| |||
7449 AGAGCAGCTGCGCGCGCAGAGAGAGCTGACGAAATTCCTGAGAGAGC 7498
      : : : : : ||| ||| : : : : : ||| |||
1321 YrAlaProProLysGlnProPheArgGlnAspValProProSerProSer 1337
      : : : : : ||| ||| : : : : : ||| |||
7499 AGCTCCCTCGCGAA.....AGGGAAGAA..... 7521
      : : : : : ||| ||| : : : : : ||| |||
1338 GlnValAlaArgLeuAsnArgLeuGlnTyrProGlnLysGlnArgProPh 1354
      : : : : : ||| ||| : : : : : ||| |||
7522 .....CAGCAGCTGCGCGCGCAGAGAGCGGAGAGAAATTT 7556
      : : : : : ||| ||| : : : : : ||| |||
1354 e 1354
7557 C 7557

```

```

seq_name: /cgm2_6/ptodata/2/1ne/5B_COMB.seq:US-08-800-644-93
seq_documentation_block:
  Sequence 93, Application US/08800644
  Patent No. 5958752
  GENERAL INFORMATION:
    APPLICANT: Steinert, Peter M.
    APPLICANT: Lee, Seung-Chul
    APPLICANT: Kim, In-Gyu
    APPLICANT: Chung, Soo-Il
    TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
    NUMBER OF SEQUENCES: 117
    CORRESPONDENCE ADDRESS:
      ADDRESSER: Knodde, Martens, Olson & Bear
      STREET: 620 Newport Center Drive, Sixteenth Floor
      CITY: Newport Beach
      STATE: CA
      COUNTRY: U.S.A.

```

```

      ZIP: 92660
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patent Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/800,644
      FILING DATE: 14-FEB-1997
      CLASSIFICATION: 424
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/056,200
      FILING DATE: 30-APR-1993
      ATTORNEY/AGENT INFORMATION:
      NAME: Fredrick, Michael F.
      REGISTRATION NUMBER: 36,799
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (714) 760-0404
      TELEFAX: (714) 760-9502
      INFORMATION FOR SEQ ID NO:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 951 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: cDNA
      HYPOTHEetical TYPE: NO
      ANTI-SENSE: NO
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 1507..1644
      FEATURE:
      NAME/KEY: Intron
      LOCATION: 1645..2511
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 2512..8070
      US-08-800-644-93

alignment_scores:
  Quality: 216.00      Length: 1219
  Ratio: 0.391        Gaps: 50
  Percent Similarity: 45.283      Percent Identity: 19.688

alignment_block:
  US-09-757-781-2 x US-08-800-644-93 ..

Align seg 1/1 to: US-08-800-644-93 from: 1 to: 951

304 LysArgLeuGlnLysGlnLysAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 320
      : : : : : ||| ||| : : : : : ||| |||
4426 GAGAGCTCGAGAGCGCGCGAGAGCGCGAGCATGAGAGAGAGAGCGCGA 4475
      : : : : : ||| ||| : : : : : ||| |||
320 uAsnAspCysIleValArgIleAsnAspLysPheLeuArgAsnArgArgP 337
      : : : : : ||| ||| : : : : : ||| |||
4476 GCAG.....GAGCTAGCTGAGAGAGC 4498
      : : : : : ||| ||| : : : : : ||| |||
337 heGlnGlnAlaGlnIleMet.PheArgGlnAlaMetArgTyrProIleT 353
      : : : : : ||| ||| : : : : : ||| |||
4499 AGGAGAGCGCGCGAGCGAGATTAGAGCGCGCATCCGAAAGTTCAGCGAGC 4548
      : : : : : ||| ||| : : : : : ||| |||
353 eTrpPheHisValAlaProAlaAlaAsnLys...GlnGlnTyrGlnGln 369
      : : : : : ||| ||| : : : : : ||| |||
4549 CAGCTGAAGAGCGAGCGCGAGCGAGCGAGCAAGCAAGAGTTCAGCGAGC 4598
      : : : : : ||| ||| : : : : : ||| |||
369 euSerGlnSerGlnLysAsnAsnTyrTyrSerArgPheSerProAsp 385
      : : : : : ||| ||| : : : : : ||| |||
4599 CCGCAGAGAGAGAGAGAGAGAGCGCGCGAGAGAGAGAGAGAGAGAGC 4648
      : : : : : ||| ||| : : : : : ||| |||
386 SerGlnTyrIle.....As 390
      : : : : : ||| ||| : : : : : ||| |||

```



```

927 .....AlaAlaIleAspLysSerTyrAspLysProAlaValAs 939
6124 TTCACGAGGAAGAACAGCTGAGCCAGAGAAAGAAAGAACGAAAG 6173
939 pAspAsp.....AspGluGluMetGluThrLeuGluGluAspThr 953
6174 GCGGACAGAGAAATCCCAAGAGAAAGCAATTCCTGAGAGAGAAAG 6223
953 Lu.....GluSerSerArgSerGlyArgGluSerVal 963
6224 AAGAGAGAGACCCCTCAAGAGACAGACAGAAATTCCTCGAGAGAA 6273
964 SerThrAlaSerAspGluProSerHisSerLeuGluArgGluMetAsn 980
6274 CACCTGCTCCAGAGAAAGGAGAGAACAGCCGCTGCTCGCCAAAGCCTGA 6323
980 Y.....AsnGluGluGlyAspLysThr 988
6324 CAGAAATTCCTCGAGAGAAAGCAATTCCTCAATCAGAAACAGAGAAAT 6373
988 hrAspArgLysAspLysThrGlyGluLysLysLysAspArgAsp 1004
6374 TCCTTCGAGAGGAACAGCGCTGCGCGAGAGAGAGAGAAATTCCTT 6423
1005 LysGluLysAspLysMetLysAlaLysGlyMetLeuLysGlyLeuGlu 1021
6424 AAGAGAGACAGCAGCTGCGCTGCGAGAGAG.....CG 6455
1021 YAspMetPheArgPheGlyLysHisArgLys...AspAspLysIleGlu 1037
6456 CGAGCAACTGCTGACAGACCGGACAGAAATTCCTCGAGAGAGAACAG 6505
1037 YThrGlyLysLysLysIleGluGluSerPheThrSerGluGluGluArg 1053
6506 AGCTGAGCCGCCAAGAGCTGACAGAAATTCCTGAGAGAGAACAGACAG 6555
1054 IleArg.....MetLysGluGluGluGluArg 1062
6556 GTGCGCGCCAGAGAGAGAGAAATTCCTGAGAGAGAGAGAGAGCT 6605
1062 gTleGlnAlaLysThrArgGluPheArgGlu..... 1072
6606 GCGCCAGAGAGGCTCAGAAATTCCTCGGAAAGAGAACAGCTGCTCAG 6655
1073 .....ArgGlnAlaArgGluArgAspLysTyrAla 1081
6656 AAGGAGAGAAACAGACAGCTGACCCGCAAGAGCGTACAGAAATTCCT 6705
1082 GluIleGlnAspPheHisArgThrPheGlyCysAspAspGluLeuMet 1098
6706 GAGGAGAGAACAGAGCTGCGCCCAAGAGCTGACAGAAATTCCTCGCA 6755
1098 rGlyGlyValSerSerTyrGluGlySerMetAla...LeuAsnAlaArg 1114
6756 ACAGAGAACTGCGCAGTACAGAGAACAGAAATTCCTCGAGAGAGAAC 6805
1114 rGlnInsProArgGluGluGlyHisMetMetAlaLeuTyrAlaGlnVal 1130
6806 AGCAGCTGACCGCCAGACGCGAGAAATTCCTCGAGAGAGAACAG 6855
1131 LysLysProArgAsnSerLysProSerProValAspSerAsnArgSer 1147
6856 CAGCTGCCCGCC.....CAGAGCGCCGCGCA 6881
1147 rProSerAsnHisAspArgIleGlnArgLeuArgGluPheGln.... 1162
6882 ACAGCGGCTGACAGCGTACAGAAATTCCTCGAGAGAGAACAGCTCG 6931
1163 ..GlnAlaLysGlnAspGluAspValGluAspArgArgThrThrLys 1178
6932 GCCAGAGAGAGGAGAGACAGACCTGACCGCCAGAGAGAGCTGACAGAA 6981
1179 PheGluGlnProTrpProAsnAlaArgProAlaThrGlnSerGlyArg 1195

```

---

```

6982 TTCGTTTAGAGAAACAGAAAGTCCGCCAGAGAACCAAGAG...AGAA 7028
1195 sSerValSerValGluValGlnMetGlnArgGln..... 1206
7029 ATTCATGAGAGACAGACAGCTGCGCCCGAGAGAGGCCCAACAGC 7078
1207 ..ArgGlnGluGluArgGluSerSerGlnGlnAlaGln..... 1218
7079 TGCCCGAGAGAGACAGAAATTCCTCGGAGAGACAGACAGCTGCTCAGAA 7128
1219 ArgGlnLysSerSerLeuProArgGlnSerArg..... 1229
7129 AGGAGAACAGCAGCTGACAGCCCAAGAGCTGACAGAAATTCCTCGA 7178
1229 ..... 1229
7179 GGAGAACCGCAGCTGCGCCGACAGAGCGGCAACACAGCTGCTCAGC 7228
1230 .....LysAsnAlaSerSerValSerGlnAspSerTrp 1240
7229 ACCGCGACAGAAATTCCTGAGAGAGAACAGCTGCTCGAGAGAGAG 7278
1241 GluGlnAsnTyrSerProGlyGluGlyPheGlnSerAlaLysGluAsnPr 1257
7279 GAAACAGCAGCTGCGCGCCGACAGAGCTGACAGAAATTCCTCGGAGAGAG 7328
1257 GArgTyrSerSerTyrGlnGlySerArgAsnGlyTyrLeuGlyGlyHisG 1274
7329 ACAGAGCTGCGCCGCTAGAGAACAGAGAGAAATTCCTC..... 7368
1274 LysPheAsnAlaArgValMetLeuGluThrGlnGluLeuArgGlnGlu 1290
7369 .....CAGAGAGAACAGCAGCTGCGCCGACAGAA 7398
1291 GlnArgArgLys.....GluGlnGlnMetLysLysGlnProProse 1304
7399 CTGAGAGAGAAATTCCTGAGAGAGAACAGCTGCGCCCAAGAGAGAG 7448
1304 rGluGlyProSerAsnTyrAspSerTyrLysValGlnAspProSer 1321
7449 AGAGCAGCTGCGCCGCAAGAGCTGACAGAAATTCCTGAGAGAGAG 7498
1321 YrAlaProProLysGlyProPheArgGlnAspValProProSerProser 1337
7499 AGCTCGCTCGGAA.....AGGAGAGAA..... 7521
1338 GlnValAlaArgLeuAsnArgLeuGlnThrProGluLysGlyArgProPh 1354
7522 .....CAGCAGCTGCGCGCCGACAGAGCTGACAGAGAAATTCCT 7556
1354 e 1354
7557 C 7557

```

seq\_name: /cgn2\_6/plodata/2/lna/5A\_COMB.seq:US-08-769-309A-4  
seq\_documentation\_block:  
; Sequence 4, Application US/08769309A  
; Patent No. 5741890  
; GENERAL INFORMATION:  
; APPLICANT: Scott, John D.,  
; APPLICANT: Nauert, Brian J.,  
; APPLICANT: Klauert, Theresa M.,  
; TITLE OF INVENTION: Protein Binding Domains of Gravin  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower/233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/769,309A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5741890and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6605 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 192..5531
US-08-769-309A-4
```

```
alignment_scores:
Quality: 207.00 Length: 1183
Ratio: 0.422 Gaps: 50
Percent Similarity: 41.420 Percent Identity: 19.104
```

alignment block:  
US-09-757-781-2 x US-08-769-309A-4 ..

Align seg 1/1 to: US-08-769-309A-4 from: 1 to: 6605

```
397 AlAGlyLeuHisThrValGlnArgAlaProAlaGluAsnHisProProgl 413
||||| ||| |||||:||||| :|||:|||||
198 GCGGAGAGCTCCACCGACGCGCAGCCG...GAGCAGCCGCCCGA 241
413 uGlnIleAspSerHisSerArgLeuProHisSerAlaHisProSerGlyL 430
| ||| ||| ||| :|||:|||||
242 G.....GGAGCTCCACGCGCGGTAGCCCGAGCCGACGCGCGC 279
430 ySProSerAlaProAlaSerAlaProGlnAsnValPheSerThrThr 446
||||| ||| ||| :|||:|||||
280 GCGGCGCCCTCGCGCGCGCGCCAGACACACCGCGACCCCGCATC 329
447 ValSerSerGlyTyrAsnThrLysLysIleGlyLysArgLeuAsnIleGl 463
:||||| ||||| :|||:|||||
330 GCGGCTCGACCCCGCCACCAAGCTCTACAGAG...AATGCTCA 373
463 nLeuLysLysGlyThrGluGlyLeuGlyPheSerIleThrSerArgAspV 480
|||||
374 GGTGTC..... 380
480 AlThrIleGlySerAlaProIleTyrValLysAsnIleLeuProArg 496
|||||:|||||
381 ..ACCATCATGCGCTAGCTGAG.....CAAGAT 407
497 GAlAlAlAlIleGlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGl 513
:|||||:|||||:|||||
408 GAGCTAGGCTCCAGAGGGTGACTTAATGCGCAGAAAGAGACCCCTGAA 457
513 uValAsnGly.....ValAspL 519
:|||||
458 CGGTCAAGAGCCCTAAACAGCAGGAGGAGAAAGAAAGATCATTTGTACGG 507
519 euValGlyLysSerGlnGluValValSerLeuLeuArgSerThrLys 535
```

```
|||||:||||| ||||| ||| |||
508 AGGTGGACAGAGAGACTGTGAAGATGTGAGCCGAAGAGACTCCGATAAA 557
536 MetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAlaPheH 552
:||||| ||| |||||
558 GAGATGGCTACTAAGTCAAGCGGTGTTCAGACATCCACAGATGATGGCCA 607
552 sProArgGluLeuAsnAlaGlu.....ProSer..... 561
:||||| ||| ||| |||||
608 GAGAGAACCGAAGAAATATCGAACAGATTCCTTTCAGAAAGCAATTTAG 657
562 ..GlnMetGlnIleProLysGluThrLysAlaGluAsp..... 573
:||||| ||| |||||:|||||
658 AAGAGTAAACAAACCCACTGATGCCAGGCTAATGATTTGATTGAAG 707
574 .....GluAspIleVal 577
708 AAGGTGTTTAAGTTTGTGGCTTAAATTCACCTGTGAAGAAAGATTAAGAC 757
577 IleuThrProAspGlyThrArgGluPheLeuThrPheGluValProLeuA 594
:||||| ||| :|||||:|||||
758 AGAGAACCTGAC...ACTGTCCAGTACTACTCTGTGAAGAAAGAT...G 801
594 snAspSerGlySerAlaGlyLeuGlyValSerValLysGlyAsnArgSer 610
:|||||:||||| |||
802 AAGGGAGGAGAGCAGCAGGGGCTGGC..... 827
611 LysGluAsnHisAlaAspLeuGlyIlePheValLysSerIleIleAsnGl 627
828 .....GACCAACAGAC.....CCACCTTGGGCTGG 856
627 yGlyAlaAlaSerLysAspGlyArgLeuArgValAsnAspIleuIleA 644
| ||||| |||
857 AGAAGCAGATCCAAA..... 872
644 lValAlaAsnGlyLysSerLeuLeuGlyLysThrAsnGlnAspAlaMetGlu 660
||||| ||| :|||||
873 .....GAAGCGAACCCAAACATCTACAGAGAAACCCGAAGAG 911
661 ThrLeuArgArgSerMetSer.....ThrGluGl 670
|||||:|||||:|||||
912 ACCCTGAAGCGTGAAGCAAGCCACGCAAAATTTCTCCCGACCGCAATC 961
670 yAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgIleSerLysC 687
: ||| :|||:|||||
962 T.....GCCAAGCAGTGAAGGAT 981
687 yAsnGlu.....Leu 690
|||||
982 GCAAGAGAGAAAGAAAGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1031
691 LysSerProGlySerProProGlyLysProGluLeu..... 701
:|||||:||||| |||
1032 GAATCTCGACTAGTCCCTGCTGACCAGTGAACAGATCAACCTTCAAAA 1081
701 ..... 701
1082 ATTCTTCACTCAAGTTGGGCGGCTGGCGCAAAAGACCGATTTCAGGA 1131
701 ..... 701
1132 AGCCGAAGAGATGAATGGAACCTTCAGAGAAAGAAAGAAAGAAAGAAAG 1181
702 .....ProIleGluThrAlaLeuAspAspArgGluArgGlyIleSerH 716
:|||||:||||| |||
1182 CCAGAAAGAAAGTACACACAGAAAGACGAAAGCAGAGGTGCTCGCA 1231
716 sSerLeuTyrSerGlyIleGlnGlyLeuAspGlyLysSerProSerArgAsnA 733
:|||||:|||||:|||||
1232 GAAGACTAGCGCTCCGAGCAAGGCCACCCACAGAGCGCGGCAAGAAAGT 1281
733 lAlaIleuSerArgIleMetGlyLysSerGlyLysTyrGlnLeu..... 747
|| |||||
```

```

1282 CCCACGAGCCCGCTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCA 1331
748 .....SerProthThrValAsnMetProGlnAs 756
1332 GAGGACAGATGATGCGCTCGAGGAGACCTTCTGAAAGAAACCTGCTCC 1381
756 pAspThrValIleIleGluAspAspArgLeuPro.....767
1382 GTTGGGACGAGAAAGTTGATGAGAAATATGAAGTCCACCAAGAAAG 1431
768 ..ValLeuProProHisLeuSer.....AspGlnSer 777
1432 TTGTGGCCGAGTCCACCTCAGCAACCTGAGAGAGAACCCGAAAGAGAC 1481
778 SerSerSerSerHisAspAspValGlyPheValThrAlaAsp.....791
1482 AAAACGGAGGTGGAAACAGCAGCGGTCTGTCCAGCTGAAAGATGGGT 1531
792 .....AlaGlyThrThrPalaIleAlaIleS 801
1532 TGGAAATGATGCAAGACCTCAGAAAGCCGAAACCTGCCAAGAGCTGTGA 1581
801 eAspSerAlaAspCys.....Ser 807
1582 AGCTCAAGAAACGTGTGTTCCGAGAGACCTTACACAGGAGACTGAC 1631
808 LeuSerProAspValAspProValLeuAlaPheGlnArgGluGlyPheG 824
1632 CTCAGTCTCGAT...GAGAAAGTCTGTCTCAAAACCCCGAAAGGGT.. 1676
824 YArgGlnSerMetSerGluLysArgThrLysGlnPheSerAspAlaSerG 841
1677 .....GTGAGTC 1683
841 ILeuAspPheValLysThrArgLysSerLysSerMetAspLeuGlyLe 857
1684 AGGTGAATGCTGTTCATCAGACGAGAGAAATGAAGTGCAG..... 1724
858 AlaAspGluThrLysLeuAsnThrValAspAspGlnLysAlaGlySerP 874
1725 .....GGAGTCC 1732
874 OSerArgAspValGlyProSerLeuGlyLeuLysLysSerSerSerLeuG 891
1733 ACTAAAGAGCTTTTACACAGACCTGCTTAAATAAGCTTTCTGGA.... 1778
891 IuSerLeuGlnThrAlaValAlaGluValThrLeuAsnGlyAspIlePro 907
1778 ..... 1778
908 PheHisArgProArgProArgIleIleArgGlyArgGlyCysAsnGluSe 924
1779 .....AMGAACGAAAGGAAAGGAGAGAGAGAGAGAGAGATC 1819
924 rPheArgAlaAlaIleAspLysSerTyrAspLysProAlaValAspAsp 941
1820 A.....GGGAGCACACTCAGGTTCTCAGCGATTCCTCGG 1854
941 sPAspGluGlyMetGluThrLeuGluAspThrGluGluSerSerArg 957
1855 ACAGCCAG.....GAGAGCAAAAGGCGGAGAGCTTCGCC 1889
958 SerGlyArgGluSerValSerThrAlaSerAspGlnProSerHisSerLe 974
1890 TCATCCCTCGAGAGACCCGAGAGAGATCAGTGTCTGAAAAAGGCTTAGC 1939
974 uGluArgIleMetAsnGlyAsnGlnGluLysGlyAspLysThrAspArg 991
1940 CGAGGTGACAGCAGATGGGAAAGCTGAAGAAGAGCTACTTCGATGAG 1989
991 ySLysAspLysThrGly..... 996
1990 AGAAAAAAGGAAGGTGTCTCCTCGGCCATCATTAAGAAAGATGTG 2039
997 ..LysGluLysLysAspArgAspLysGluLysAspLysMetLysAl 1012
2040 ACGCCCAAGAGAGCTGTAGACGCGCTTGGAAAGTAT..... 2078
1012 aLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArgPheGlyLysH 1029
2078 ..... 2078
1029 IAspGlyAspAspLysIleGlu.....LysThrGlyLysIleLysIle 1043
2079 .....AAGAAGATGAGCTCGACAAAGTCAAGAGCGCTACCTGTCTTCC 2123
1044 GlnGluSerPheThrSerGlu...GluGluArgIleArgMetLysGlnG 1059
2124 ACCGAGAGCACAGCTCTGAAATGCAAGAAATGAAGAAAGGAGCGTGA 2173
1059 uGlnGluArgIleGlnAlaLysThrArgGluPhe..... 1070
2174 AGAGCCAAAGCCGGAAGAAACCAAGCGAGGTGATACCTAGTATCTT 2223
1071 .....ArgGluArgGlnAlaArgGluArg 1078
2224 GGAAGCTTTAATTTGTGTGGATCATCCAAAGAAAGCAGAGAGAG 2273
1079 AspTyrAlaGluIleGlnAspPheHisArgThrPheGlyCysAsp.... 1093
2274 TCCTCTTGATGAGAAAGGAGCCAAAGCAATGAGAGAGACACCA 2323
1094 .....AspGluLeuMetLysGlyLysValSerSerTyrGluGlySerM 1108
2324 GAAAGCTGATGAGCCGGAAGAAACAAAGAGAGCGGAGACAGCGGATCC 2373
1108 eAlaLeuAsnAlaArgProGlnSerProArgGluLysIleMetLys 1124
2374 TTGCT...GGTCCCAAGAACATGATCCAGCGAGGA..... 2408
1125 AlaLeuTyrAlaGlnValLysLysProArgAsnSerLysPro..... 1138
2409 .....AGTTCCTCCCGGAGCAAGC 2428
1139 .....SerProValAspSerAsnArg...SerThrProSerAsnHisAsp 1153
2429 TGGAAACCTTACCGAAGGGGAGCGGCTTCCACCTGGAGATCATTTAAA 2478
1153 rGlnGluArgLeuArgGlnGluPheGlnGlnAlaLysGlnAspGluAsp 1169
2479 GGTACTCAGCCCAAGAAAAA...TCAAGTCCAAAGCTGGAAGAGAA 2525
1170 ValGluAspArgArgArgThrTyrSerPheGluGlnProThrProAsnAl 1186
2526 AGCGAAAGACTCCATAGCTGGGTCTGTGTAGACATTCATCCATCAGAC 2575
1186 aArgProAlaThrGlnSerGlyArgHisSerValSerValGluValGln 1203
2576 TGAACCGGTAAAGAAAGATCTGGGTCTCAATCAAGAAAGTTATTCCG 2625
1203 eGlnArgGlnArgGlnGluGluArgGluSerGln..... 1215
2626 GACGAGAGAGAAAGGCCAGATGGGAACAAAGCAACCCCTGTGAA 2675
1215 ..... 1215
2676 GAGCAGAGGCCAACAGGGGCCAACGAAGATGATGTGATGCCGGCGT 2725
1216 .....GlnAlaG 1218
2726 GGTCCCTGTGTGATATGATGCTGTGAAGAGGAGAAATGGAGGAC 2775
1218 IAspGlnTyrSerSerLeuProArgGlnSerArgLysAsnAlaSerSer 1234
2776 AGCAAGCCCAAGAGCGCAGAGAGCCGAGCAAGAAAGCAGCACATGAG 2825

```







OM of: US-09-757-781-2 to: N\_Geneseq\_032802:\* out\_format : pfs

Date: Jul 24, 2002 4:27 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

#### Command line parameters:

-MODEL=framed+p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09757781/runit.22072002.154235.17679/app\_query.fasta\_1.1434  
-DB=N\_Geneseq\_032802 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blissum62  
-TRANS=human40.cdl -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORTEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09757781@cgn2.1.238 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPAY -WAIT -THREADS=1

#### Search information block:

Query: US-09-757-781-2  
Query Length: 1356  
Database: N\_Geneseq\_032802:\*  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 322.720000

#### score\_list:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation	..	7676.07	0.0
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH73337								6480.00	5222.18
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH18587								4409.00	5222.18
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH78369								4409.00	5222.18
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH01182								3527.00	4174.83
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH27293								2778.50	3288.85
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH27320								1397.00	1654.91
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH33594								1397.00	1654.91
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH360861								1187.50	1397.02
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH56056								1187.50	1397.02
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH60963								1187.50	1397.02
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH61007								1187.50	1397.02
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH08765								1171.50	1368.95
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH08765								942.00	1078.02
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH07109								936.00	1110.93
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH71480								902.50	1070.45
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAH79284								852.50	1012.00
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH23370								852.50	1012.00
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH68902								784.00	931.16
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH67583								694.00	825.97
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH70055								675.00	801.77
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH80567								671.00	799.05
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH12111								621.50	737.12
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH08393								603.50	709.75
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH66903								590.00	691.98
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH66507								528.00	629.32
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH72113								472.50	565.98
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH71048								472.50	565.98
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH69243								459.50	549.52
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH68901								433.50	513.45
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH29911								377.50	423.55
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAH47414								355.50	404.48
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAH49920								355.50	397.29
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAH36453								355.50	396.37
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAH29908								352.00	396.24
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAH29909								351.00	392.25
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAH29960								343.50	391.40
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAH6366								340.50	380.86
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH71475								338.00	379.82
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH51926								338.00	379.74
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH71477								338.00	379.64

/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH70304 + 338.00 379.39 8.7  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH70303 + 338.00 378.73 9.5  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH71476 + 338.00 378.73 9.5  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH71475 + 338.00 378.25 1.0  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAH85924 + 332.00 367.69 3.96

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH73337

#### seq\_documentation\_block:

ID AAH73337 standard; cDNA; 5510 BP.

XX AAH73337;

AC 19-SEP-2001 (first entry)

DE Human cervical cancer marker nucleic acid 4611.

KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX Homo sapiens.

PN WC200142467-A2.

PD 14-JUN-2001.

XX 08-DEC-2000; 2000WC-US33312.

PF 08-DEC-1999; 99US-0169681.

PR 21-DEC-1999; 99US-0171350.

PR 14-MAR-2000; 2000US-0189315.

PR 12-MAY-2000; 2000US-0203791.

PR 09-JUN-2000; 2000US-0210600.

PR 21-JUL-2000; 2000US-0220114.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Deeds J, Berger A, Zhao X;

XX WPI; 2001-375006/39.

PT New isolated nucleic acid for diagnosing and treating cervical cancer

PT and for assessing and detecting compounds for treating the cancer -

XX Claim 1; Page 1029-1031; 1051pp; English.

XX The invention relates to novel genes (AAH68727-AAH73383) associated with

XX polypeptides are useful: to assess if a patient is afflicted with

XX cervical cancer or has a pre-malignant condition; to monitor the

XX progression of cervical cancer or a premalignant condition in a patient;

XX and to select and/or assess the efficacy of a compound or therapy for

XX inhibiting cervical cancer in a patient. The nucleic acids may also be

XX useful for gene therapy.

SQ Sequence 5510 BP; 1526 A; 1242 C; 1418 G; 1305 T; 19 other;

alignment\_scores: Quality: 6480.00 Length: 1355

Percent Similarity: 93.358 Percent Identity: 93.358

alignment\_block: US-09-757-781-2 x AAH73337 ..

Align seg 1/1 to: AAH73337 from: 1 to: 5510

2 LysValThrValCysPheGlyArgThrArgValValProCysGlyAs 18

||||| 4 AAATGACCGTGTCTTGGACGGACCGGGTGTGTGCTGCGTGGGGA 53

18 pGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgTyr 35

|||||

54 CGGCACATGAAAGTTTTCAGCCTCATCCAGCAGCGGCTGACCCGCTACC 103  
35 rGlySalIaIleAlaLysAspProAsnTyrTrpIleGlnValHisArgIleu 51  
|||||  
104 GGAAGGCCATCGCCAAAGATCCAAACTACTGATACAGGTGCATCGCTTG 153  
52 GlnHisGlyAspGlyIylIleuAspLeuAspAspIleuGlyAspA 68  
|||||  
154 GAACATGAGATGAGGAATACTAGACCTTGATGACATTCTTTGTATGT 203  
68 lAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAspPro 85  
|||||  
204 AGCAGACGATAAAGACACCTGGTACAGGTGGTTGATGAGCAGATCCAC 253  
85 lHisIleGlyLysAspGlyThrSerAlaSerSerThrGlyThrGlnSerPro 101  
|||||  
254 ATCAGCGAGGTGATGCGCACCAAGTCCACGGGTACCCGAGACCCCA 303  
102 GluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheGlnPr 118  
|||||  
304 GAGATATTGGTAGTGAAGCTGGCACCAACAATGCTCAGCCTTTCAGCC 353  
118 oTyrGlnAlaThrSerGluIleGlnValThrProSerValIleuArgAla 135  
|||||  
354 TTACCAAGCAACAAGTGAATGAGGTCACTCAGTCCCTTCGAGCAA 403  
135 sMetProLeuHisValArgArgSerSerAspProAlaLeuIleGlyLeu 151  
|||||  
404 ATATGCTCTTTCATGTTGACGCGAGTACGACCCAGCTCTAATTGGCC 453  
152 SerThrSerValSerAspSerAsnPheSerSerGluGlnProSerArgly 168  
|||||  
454 TCCACTTGTGTCAGTGTAGTAAATTTTCTCTGAAAGACCTTCAAGAA 503  
168 sAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnThrA 185  
|||||  
504 AATCCCAACACCTGGTCAACAACACCTGGCTCTCAAGCAACAACAG 553  
185 lAgLysErProLysThrCysAspArgLysLysAspGluAsnTyrArgSer 201  
|||||  
554 CTGGGAGCTCTTAAACCTGCGACAGGAG..... 582  
202 LeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAsnAl 218  
582 ..... 582  
218 aArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuGln 235  
582 ..... 582  
235 yGlnGlnGlnAspGluAspGlyThrGlnGluAspAsnSerArgValGln 251  
|||||  
583 ..... GATGAGATGGGAGAGAGAGATTAACAGTGTGTGA 621  
252 ProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSerIe 268  
|||||  
622 CCGTGGACATGCTGACACGGGTTGGAGCATATACCAACTTTTCTCT 671  
268 uAspAspMetValLysLeuValGluValProAsnAspGlyGlyProLeuG 285  
|||||  
672 GATGATATATGATTAACCTGTAGAAAGTCCCAACATGAGAGGCTCTG 721  
285 lYlIleHisValValProPheSerAlaArgGlyLysArgThrLeuGlyLeu 301  
|||||  
722 GAATCATATAGTACCTTTCAGTGTCCGAGGCGGCAAACTGGGGTTA 771  
302 LeuValLysArgLeuGlnLysGlyLysAlaGlnHisGlnAsnLeuPh 318  
|||||  
772 TTAGTAAACGATGAGAAAGTGTAAAGCTGAACATGAAAACTTTT 821  
318 eArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArgAsnA 335  
|||||  
822 TCGTGAAGAAATGATTCATGTCAGGATTAATGATGCGCACCTTCGAATA 871

335 rGArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThrPro 351  
|||||  
872 GAAATTTTGAACAAGCACACATATATTTTCGCCAACCCATGCTACACCC 921  
352 lIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnTyrGluG 368  
|||||  
922 ATCATTTGGTCCATGTGGTTCCTGCAGCAAAATAAAGACAGATGACAA 971  
368 nLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerAlaArgPheSerPro 385  
|||||  
972 ACTATCCCAAGGAGAAACAATTAATTAATTCACCCCTTTTAAAGCCCTG 1021  
385 sPserGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHisThr 401  
|||||  
1022 ACAGCCAGATATATGCAACAGAGGTGTACACAGTCCAGGCTTTCACAG 1071  
402 ValGlnArgAlaProArgLeuAsnHisProGluGlnIleAspSerH 418  
|||||  
1072 GTCCAGAGAGCAACCCGACTGAACCCCGCTGAGCAGATGACTTCGA 1121  
418 sSerArgLeuProHisSerAlaHisProSerGlyLysProProSerAlaP 435  
|||||  
1122 CTCGAAGACTACCTCATGAGCGCACACCCCTCGGGAACAACCATCCGCTC 1171  
435 roAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGlyTyr 451  
|||||  
1172 CAGCGTCGGCACCTCGAATGTATTTAGTACAGCTAAGCAGTGGTTAT 1221  
452 AsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysGlyTh 468  
|||||  
1222 AACACCAAAAAAATAGCAAGAGGCTTAATATCCACCTTAAGAAAGGTAC 1271  
468 rGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyLys 485  
|||||  
1272 AGAAGTTTGGATTTCAGATCATCTTCAGAGATGTAACAATAGTGCT 1321  
485 eAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIleGln 501  
|||||  
1322 CAGCTCATATGATGAAAAACATTTCTCCCGGGGGCGGCCCATTCAG 1371  
502 AspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyValAs 518  
|||||  
1372 GATGGCCGACTTAAGCGACGAGACACACTTAATAGAGTAAATGAGTAGA 1421  
518 rLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSerThrL 535  
|||||  
1422 TTTAGTGGCAAAATCCCAAGAGAAAGTTGTTCGCTTGAGAAAGCACCA 1471  
535 yMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAlaPhe 551  
|||||  
1472 AATGGAAGAACTGTGAGCCTTGTGCTTTCGCCAGAAACCGCTTCG 1521  
552 HisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLysG 568  
|||||  
1522 CACCCAAAGGAACTG..... 1536  
568 uThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrArg 585  
|||||  
1537 .... AAGCAGAAAGATGAGGATATTTCTTAACCTGATGGCACCAAGG 1582  
585 lUpHeLeuThrPheGluValProLeuAsnAspSerGlySerAlaGlyLeu 601  
|||||  
1583 AATTCTGTGATTTGAAGTCCACATTAATGATTCAGGATCTCAGACCT 1632  
602 GlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLeuG 618  
|||||  
1633 GGTGTCAAGTCAAAAGGTAAACGGTCAAAAGCAACACAGCAGATTGGG 1682  
618 yLlePheValLysSerIleIleAsnGlyLysAlaAlaSerLysAspGlyA 635  
|||||  
1683 AATCTTTGTCAAGTCCATTATTAATGAGAGACACATCTTAAGATGGA 1732

```

635 rgleuArYValAsnAspGlnLeuIleAlaValAsnGlySerLeu 651
|||||
1733 GGCTTGGGGATGATCAACTGATGACGATTAATGAGATCCCTGTT 1782
652 GlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSerTh 668
|||||
1783 GCGAACAACAACCAAGATGCCATGGAAACCCCTAAGAAAGCTATGCTAC 1832
668 rGluGlyAsnLysArgGlyMetIleGlnLeuIleAlaAlaArgArgIle 685
|||||
1833 TGAAGCAATTAACGAGAGATGATCCAGCTTATTGTTGCAAGAGAAATA 1882
685 erLysCysAsnGlnLeuLysSerProGlySerProGlyProGluLeu 701
|||||
1883 GCAAGTCGATAGCTGAAGTCACTGGGAGCCCCCTGGACCTGAGCTG 1932
702 ProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSerI 718
|||||
1933 CCCATTGAAACACGCTTGATGATGAGAGAGAGAAATTTCCCATTTCC 1982
718 uTYrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAlaAla 735
|||||
1983 CTACAGTGGGATGAGGGGCTTGATGATCCCAAGCAGAAATGCTGCC 2032
735 euSerArgIleMetGlyGluSerGlyLysTYrGlnLeuSerProThrVal 751
|||||
2033 TCAGTGGGATGAATG.....GGTAAATACAGCTGTCCCTACAGTGT 2073
752 AsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuProVa 768
|||||
2074 AATATGCCCCAAGATGACACTGTCATTATAGAAAGATGACAGGTGGCAG 2123
768 IleuProThrHisLeuSerAspGlnSerSerSerSerHisAspAspV 785
|||||
2124 GCTTCTCCACATCTCTGACAGTCCCTCTTCAGCTCCCATGATGATG 2173
785 aGlyLysHeValThrAlaAspAlaGlyThrTYrPalalysAlaAlaIle 801
|||||
2174 TGGGGTTTGGAGCGGAGATGTGGTACTTGGGCCAAGGCTGCATACGT 2223
802 AspSerAlaAspCysSerLeuSerProAspValAspProValLeuAlaP 818
|||||
2224 GATTACGCCACATGCTTTTGAGTCCAGATGTTGATTCATTTCTGCTTT 2273
818 eGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgTYrLysG 835
|||||
2274 TCAACGAGGAAGATTGGACGTCAG..... 2298
835 InPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSerLys 851
2298 ..... 2298
852 SerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAspAs 868
|||||
2299 .....ATAGCTGACGAGACTTAACCTCAATACAGTGGATGA 2333
868 pGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyLeu 885
|||||
2334 CCAAGAAAGCGAGTTCTCCAGAGAGATGGGTCCTTCCCTGGGCTGTA 2383
885 yLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluValThr 901
|||||
2384 AGAAGTCACACTGTTGGAGAGCTTCGACAGCCGAGTTGCCGAGGTACT 2433
902 LeuAsnGlyAspIleProPheHisArgProArgProArgIleIleAspG 918
|||||
2434 TTGATGCGGATATTCCTTCCATCGTCCACGCGCGCGATTAATCACAAG 2483
918 yArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerTYrAspL 935
|||||
2484 CAGGGGATGCAATGAGAGCTTCAGAGCTGCCATGACAAATCTTATGATA 2533
935 ySProAlaValAspAspAspAspGlnLysMetGluThrLeuGluGluAsp 951
|||||
2534 AACCCCGGTAGATGATGATGAAGGACATGGAGCTTGGAGGAAGAC, 2583
952 ThrGluGlnSerSerArgSerGlyArgGluSerValSerThrAlaSerAs 968
|||||
2584 ACAAGAAAGAAAGTTCAAGATCAGGAGAGAGACTCTTATCCACAGCAGTA 2633
968 pGlnProSerHisSerLeuGluArgGlnMetAsnGlyAsnGlnGluLysG 985
|||||
2634 TCAGCCTTCCACCTCTCTGGAGAGACAAATGAATGAAACCAAGAGAAAG 2683
985 LysPlySThrAspArgLysLysAspLysThrGlyLysGluLysLysLys 1001
|||||
2684 GTGATTAAGCTGATGAGAAAAAGATAAACTGGAAAAAGAAAAAGAAAA 2733
1002 AspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLeuL 1018
|||||
2734 GATRAGATTAAGCAGAAAGCATTAATTAAGCCCAAGAAAGGAAATGCTGA 2783
1018 sGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAspLysI 1035
|||||
2784 GGGCTTGGAGACATGTTCAAGGTTGGCAAAACATCGAAAGATGACAAGA 2833
1035 leGluLysThrGlyLysIleLysIleGlnGluSerPheThrSerGluGlu 1051
|||||
2834 TTGAGAAACGGGTAAATAAATAACAGAGATCCTTACATCAGAAAGAG 2883
1052 GluArgIleArgMetLysGlnGluGlnGluArgIleGlnAlaLysThrAr 1068
|||||
2884 GACAGGATACGATTAAGCAAGCAGACAGCAGAGATTCACGCCAANAATCG 2933
1068 gGluPheArgGluArgGlnAlaArgGluArgAspTYrAlaGluIleGlnA 1085
|||||
2934 AGATTTAGGGAAGACAGACAGCTCGAGAGCGTACATAGTCAATTCAG 2983
1085 sPheHisArgThrPheGlyCysAspAspGluLeuMetTYrGlyGlyVal 1101
|||||
2984 ATTTTCATCGGACATTTGGCTGTGATGATGAGTTAATGTAGGGGAGTT 3033
1102 SerSerTYrGluGlySerMetAlaLeuAsnAlaArgProGlnSerProAr 1118
|||||
3034 TCTTCTTATGAAGTTCCATGCGCTCTCAACGCTAGACCTCAGAGCCACG 3083
1118 gGluGlyHisMetAspAlaLeuTYrAlaGlnValLysLysProArg 1135
|||||
3084 AGAAGGCAATGATGATGATGCTTGTATGCCCAAGTCAAGAACCGCGGA 3133
1135 snSerLysProSerProValAspSerAsnArgSerThrProSerAsnHis 1151
|||||
3134 ATTCCAAACCTTCACCTGTAGACAGTAAACAGATCAACTCTACGCAATCAT 3183
1152 AspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAspG 1168
|||||
3184 GATCGGATACAGCCTCTGAGGCAAGAAATTTACAGACCAAGCAAGATGA 3233
1168 uAspValGluAspArgArgArgThrTYrSerPheGluGlnProTYrProA 1185
|||||
3234 AGATGTAGAAAGATGTCGGCGGACCTATAGTTTGAGCAAAACCTGCGGA 3283
1185 snAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGluVal 1201
|||||
3284 ACGCAGCGCCGCGCAGAGAGCGGCGACACTGCTGCTCGTGAGAGGTG 3333
1202 GlnMetGlnArgGlnArgGlnGluGluArgLysSerSerGlnGlnAlaG 1218
|||||
3334 CAGATGCAAGCGCGCAGCGCAGAGAGAGCGCAGAGCTCCACGACGCCCA 3383
1218 nArgGlnTYrSerSerLeuProArgGlnSerArgLysAsnAlaSerSerV 1235
|||||
3384 GCGCGAGTACAGCTCTGCTCGCTCGCAAAAGCAGAAAAAATCCAGCTCGG 3433
1235 aLsErGlnAspSerTYrGluGlnAsnTYrSerProGlyGluGlyPheGln 1251
|||||

```

```

3434 TCTCCAGACTCTTGGAGGACGAAGTACTCCCTGGGGAAGGCTTCAG 3483
1252 SerAlaValAsnProArgTyrSerSerTyrGlnGlySerArgAsnGln 1268
3484 AGTGGCAAGAGAACCCAGGACTCCAGCTACCAAGGCTCCAGAAACGG 3533
1268 yTyrLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGlnThrGln 1285
3534 CTACTGGGAGGACATGGCTTCAACGCGAGGTATGCTGGAAACTCAGG 3583
1285 LeuLeuArgGlnGlnGlnArgArgGlyGlnGlnMetLeuValGln 1301
3584 AGCTCTTGGCCCGAGAACAGAGCGGAGAGACACACATGTAAGAACAG 3633
1302 ProProSerGlnGlyProSerAsnTyrAspSerTyrTyrLeuValGlnAs 1318
3634 CTGCTCTCCAGAGGGGCCAGCACTATGACTGCTATTAACAAAGTCCAG 3683
1318 ProSerTyrAlaProProLysGlyProPheArgGlnAspValProProS 1335
3684 CCCAGTACGCGCCCTCCCAAGGCGCTCCCGCAAGATGCGCCCTCCT 3733
1335 eProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGlnLysGly 1351
3734 CCCCCTTCTCAGGTTCGAGAGGTGAACAGACTCAGACTCCTGAGAAAGG 3783
1352 ArgProPheTyrSer 1356
3784 AGGCCCTTCTATATTC 3798

```

seq\_name: /SIS1/gcgdata/hold-1/geneseq/geneseq-emb1/NA2001A.DAT:AAH18587

seq\_documentation\_block:

ID AAH18587 standard; cDNA; 2718 BP.

AAH18587:

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:18774.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000BP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu N, Nagai K, Otsuki T;

WPI: 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602

full-length cDNAs defined in the specification, and for the detection

of an oligonucleotide comprising a sequence complementary to a

sequence and an oligonucleotide comprising a 3'-end sequence, where the

comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX Sequence 2718 BP; 840 A; 586 C; 688 G; 604 T; 0 other;

alignment\_scores: Length: 909  
Quality: 4409.00 Gaps: 2  
Ratio: 5.068  
Percent Similarity: 95.710 Percent Identity: 94.829

alignment\_block:  
US-09-757-781-2 x AAH18587 ..

Align seg 1/1 to: AAH18587 from: 1 to: 2718

```

126 GluValThrProSerValLeuArgAlaAsnMetProLeuHisValArgAr 142
3 GAGGTCAACACCTTCAGTCTTCGAGCAAAATATGCTTATATTTGAGAG 52
142 gSerSerAspProAlaLeuIleGlyLeuSerThrSerValSerAspSera 159
53 CAGTGTGACCCAGCTTAATTGGCCTTCACATTGTCAGTGAATGTA 102
159 snPheSerSerGlnGluProSerArgLysAsnProThrArgTyrSerThr 175
103 ATTTTTCCTGTGAGAGAGCCCTTCAGAAATAATCCACACGCTGTCAACA 152
176 ThrAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysThrCysAs 192
153 ACAGCTGGCTTCTCAAGCAGAAACACTGCTGGAGTCTTAAGCCTGCGA 202
192 PArgLysLysAspGluAsnTyrArgSerLeuProArgAspThrSerAsn 209
203 CAGGAAGAAAGATGAAACTACAGAAAGCCTCCGCGGATACTAGTAACT 252
209 rPheSerAsnGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSer 225
253 GGTCTTAACCAATTTCAGAGACAAATGCTCGCTGCTGAGTCCAGT 302
226 HisProMetValGlyLysTyrLeuGlnLysGlnGlnLysAspGluAspGln 242
303 CACCAATGGTGGGCAAGTGGCAGAGAAACAAAGAACAGATGAGGATGG 352
242 yThrGluGluAspAsnSerArgValGluProValGlyHisAlaAspThrG 259
353 GACAGAAAGAGATACAGTCTGTGTAACCTGTTGGACATGTCGACACGG 402
259 yLeuGlnHisIleProAsnPheSerLeuAspAspMetValLysLeuVal 275
403 GTTTGGACATATACCCCAACTTTCTCTGATGATGATGATGATGATGAT 452
276 GluValProAsnAspGlyGlyProLeuGlyTyrIleHisValValProPhe 292
453 GAAGTCCCAAGATGAGAGGCTCTGGAATTCATGATGATGATGATGATGAT 502

```

292 rAlaAaGgLyGlyArgThrLeuGlyLeuValLysArgLeuGluysG 309  
 |||:::|||||  
 503 TGCTCAGGCGGCGAACCCTGGGGTTATTAGTAAAAAGATTGGAGAAAG 552  
 309 LyGlyLysAlaGluHisGluAsnLeuPheArgGluAsnAspCysIleVal 325  
 |||:::|||||  
 553 GTGGTAAAGCTGAACATGAAAATCTTTTGTGTGAGAAATGATTCATTTGTC 602  
 326 ArgIleAsnAspGlyAspLeuArgAsnArgArgPheGluGlnAlaGlnH 342  
 |||:::|||||  
 603 AGGATTATATATGCGACCTTCGAAATAGAAATTTGAACAGACACACA 652  
 342 sMetPheArgGlnAlaMetArgThrProIleIleTrpPheHisValValP 359  
 |||:::|||||  
 653 TATGTTTCGCCAAGCCATGCGTACACCATCATTTGGTTCATGTGGTTC 702  
 359 rAlaAlaAsnLysGluGlnTrpGluGlnLeuSerGlnSerGluLysAsn 375  
 |||:::|||||  
 703 CTGCAGCAAAATTAAGACAGTATGAACTATCCCAAGTGAAGAAAGAC 752  
 376 AsnTrpTrpSerSerArgPheSerProAspSerGlnTrpIleAspAsn 392  
 |||:::|||||  
 753 AATTACTATTCAGCCGCTTTAGCCCTGACAGCCAGTATATTGACAAACAG 802  
 392 gSerValAsnSerAlaGlyLeuHisThrValGlnArgAlaProArgLeu 409  
 |||:::|||||  
 803 GAGTGTGAACAGTGCAGGCGCTCACACGCTGCAGAGACACCCGACGTGA 852  
 409 snHisProProGluGlnIleAspSerHisSerArgLeuProHisSerAla 425  
 |||:::|||||  
 853 ACCACCGCTGAGCAGATAGACTCTACTCAAGACTCACTCAATGAGGCA 902  
 426 HisProSerGlyLysProProSerAlaProAlaSerAlaProGlnAsn 442  
 |||:::|||||  
 903 CACCCCTCGGGAAACACCATCGCTCCAGCCTCGGCACCTCAGATGT 952  
 442 lPheSerThrThrValSerSerGlyTrpAsnThrLysLysIleGlyLys 459  
 |||:::|||||  
 953 ATTAGTACACACTGTAAAGCAGTGTATTACACCAAAAAAATAGCGCAAG 1002  
 459 rGleuAsnIleGlnLeuLysGlyThrGluGlyLeuGlyPheSerIle 475  
 |||:::|||||  
 1003 GGCTTATATTCAGCTTAAGAAAGTACAGAAAGTTTGGGATTCAGCATC 1052  
 476 ThrSerArgAspValThrIleGlyLysSerAlaProIleTrpValLysAs 492  
 |||:::|||||  
 1053 ACTTCCAGAGATGTAAACAATAGTGGCTCAGCTCCATCTATGTGAAAA 1102  
 492 nIleLeuProArgGlyAlaAlaIleGlnAspGlyArgLeuLysAlaGly 509  
 |||:::|||||  
 1103 CATTCCTCCCGGGGGCGGCCCATTCAGATGCCGACTTAAGCAGACAG 1152  
 509 sPArgLeuIleGluValAsnGlyValAspLeuValGlyLysSerGlnGlu 525  
 |||:::|||||  
 1153 ACAGACTTATAGAGTAAATGAGTATTAAGTGGCAAAATCCCAAGAG 1202  
 526 GluValValSerLeuLeuArgSerThrLysMetGluGlyThrValSerI 542  
 |||:::|||||  
 1203 GAAGTGTTCGCTGTGAGAGACCAACAATGGAAGAACTGTGACCT 1252  
 542 uLeuValPheArgGlnGluAspAlaPheHisProArgGluLeuAsnAlaG 559  
 |||:::|||||  
 1253 TCTGGTCTTTCGCCAGAGACGCTTCACCCAAAGGAACTGAAATCAG 1302  
 559 lProSerGlnMetGlnIleProLysGluThrLysAlaGluAspGluAsp 575  
 |||:::|||||  
 1303 AGCCAAAGCCAGATGCAGATTCCAAAGAAAGCAAGAAAGATGAGAGAT 1352  
 576 IleValLeuThrProAspGlyThrArgGluPheLeuThrPheGluValP 592  
 |||:::|||||  
 1353 ATTGTCTTACACCTGATGGCACACAGGGAATTTCTGACATTGTGAAGTCCC 1402

592 OlEuAsnAspSerGlySerAlaGlyLeuGlyValSerValLysGlyAsn 609  
 |||:::|||||  
 1403 ACTTACGTATTCAGATCTGCAGGCTTGGTGTGATGCTCAAGAGTAACC 1452  
 609 rGSerLysGluAsnHisAlaAspLeuGlyIlePheValLysSerIleIle 625  
 |||:::|||||  
 1453 GGTCAAAAGAGAACACGAGATTGGGAATCTTTGTCAAGTCCATTAAT 1502  
 626 AsnGlyGlyAlaAlaSerLysAspGlyArgLeuArgValAsnAspGlnLe 642  
 |||:::|||||  
 1503 AATGAGAGACAGCATCTTAAAGATGGAAGCTTGGGTGAATGTCAACT 1552  
 642 uIleAlaValAsnGlyLeuSerLeuLeuGlyLysThrAsnGlnAspAla 659  
 |||:::|||||  
 1553 GATAGCAGTAAATGAGAAATCCCTGTGGGCAAGACAAACCAAGATGCCA 1602  
 659 etGluThrLeuArgArgSerMetSerThrGluGlyAsnLysArgGlyMet 675  
 |||:::|||||  
 1603 TGGAAACCTTAAGAAAGCTATGTCTACTGAAGGCAATAAACGAGGAATG 1652  
 676 IleGlnLeuIleValAlaArgArgIleSerLysCysAsnGluLeuLysSe 692  
 |||:::|||||  
 1653 ATCCAGCTTATTTGTTGCCAAGAGAAATGAAGCAAGTCAATGACGTGAAGTC 1702  
 692 rProGlySerProProGlyProGluLeuProIleGluThrAlaLeuAsp 709  
 |||:::|||||  
 1703 ACCTGGAGAGCCCCCTGACCTGAGCTGCCATGAAACACGTTGGATG 1752  
 709 sPArgGluArgArgIleSerHisSerLeuTrpSerGlyIleGluGlyLeu 725  
 |||:::|||||  
 1753 ATGAGAAACGAAGAAATTTCCATTCCCTTACAGTGGGATGGAAGGGCTT 1802  
 726 AspLysSerProSerArgAsnAlaAlaLeuSerArgIleMetGlyLysSe 742  
 |||:::|||||  
 1803 GATGAATCGCCACAGCAAGATGCTGCCCTCAGTAGAATAAG..... 1844  
 742 rGlyLysTrpGlnLeuSerProThrValAsnMetProGlnAspAspThrV 759  
 |||:::|||||  
 1845 .GGTAAATACACAGCTGTCCCTACAGTGAATATCCCCAAAGATGACACTG 1893  
 759 aIleIleGluAspAspArgLeuProValLeuProProHisLeuSerAsp 775  
 |||:::|||||  
 1894 TCATTATAGAAATGACAGGTGCGCAGGTCTTCCTCCACATCTCTCTAC 1943  
 776 GlnSerSerSerSerHisAspAspValGlyPheValThrAlaAspAl 792  
 |||:::|||||  
 1944 CAGTCCCTCTCCAGCTCCCATGATGATGTGGGTTTGTGAGCGCAGATGC 1993  
 792 aGlyThrTrpAlaLysAlaAlaIleSerAspSerAlaAspCysSerLeu 809  
 |||:::|||||  
 1994 TGTACTCTGGGCCCAAGGCTGCATCAGATTCAGCCGACGTCTCTTGA 2043  
 809 eTrProAspValAspProValLeuAlaPheGlnArgGluGlyPheGlyArg 825  
 |||:::|||||  
 2044 GTCCAGATGTGATCCAGTCTTCTGTTTCAACGAAAGAAATTTGGAGT 2093  
 826 GlnSerMetSerGluLysArgThrLysGlnPheSerAspAlaSerGlnLe 842  
 |||:::|||||  
 2094 CAGACT..... 2099  
 842 uAspPheValLysThrArgLysSerLysSerMetAspLeuGlyIleAla 859  
 |||:::|||||  
 2100 .....G 2100  
 859 sPGLuThrLysLeuAsnThrValAspAspGluLysAlaGlySerProSer 875  
 |||:::|||||  
 2101 ACAGAGACTAAATCAATACAGTGGATGACCGAAAGCGGTTCTCCACGC 2150  
 876 ArgAspValGlyProSerLeuGlyLeuLysLysSerSerSerLeuGluSe 892  
 |||:::|||||  
 2151 AGAGATGTGGGTCTCTCCCTGGGTCTGAAGAAGTCAACCTATTGTGAGAG 2200  
 892 rLeuGlnThrAlaValAlaGluValThrLeuAsnGlyAspIleProPhe 909

```

|||||
2201 TCTGAGACCGCATGTCGAGGTACTTAAATGGGATTCCTTCC 2250
909 1AARGPRoARgPRoARg1le1learg1yarg1ygcysasgluserPhe 925
2251 ATCGTCCAGCGCCGGGATTAATCAGAGGAGGGGATGCAATGAGAGCTTC 2300
926 ARGAlAAl1leasplYserTyrAspLysProAlaValaAspAspAs 942
2301 AGAGGTGCCATCCACAAATCTTATGATTAACCCGGGTAGATGATGATGA 2350
942 pGIuGIuMetGIuThrLeuGIuGIuAspThrGIuGIuSerArqSerg 959
2351 TGAAGGATGAGACCTTGGAAGAACACAGAAAGAAATGTAAGATCAG 2400
959 1yargGIuSerValSerThrAlaSerAspGIuProSerHisSerLeuGIu 975
2401 GGAGAGGTGTGTATCCACAGCCAGTATCAGGCTTCCACATCTCTGGAG 2450
976 ArgGIuMetAsnGIuAsnGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 992
2451 AGACAAATGATGGAACCAAGAAAGGTGATTAAGCTGATGAAAAA 2500
992 sAspLysThrGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 1009
2501 GGATTAACCTGGAAAGAAAAAGAAAGATAGATTAAGAGAGAGATA 2550
1009 ySMetLysAlaLysLysGIuMetLeuLysGIuLeuGIuAspMetPhearq 1025
2551 AAATGAAGCCAGAGAGGAGATGCTGAAGGCTTGGAGACATGTTACG 2600
1026 PheGIuLysHisArgLysAspAspLys 1034
2601 CTTCGCAAACTGAAAGCCGAGAGAGCA 2627

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH78369
seq_documentation_block:
ID AAH78369 standard; DNA: 2718 BP.
XX
XX AAH78369;
XX
XX 26-NOV-2001 (first entry)
XX
XX Nucleotide sequence of a human protein kinase/protein phosphatase.
XX
XX Human; protein kinase; protein phosphatase; signal transduction; ss.
XX
XX Homo sapiens.
XX
XX key Location/Qualifiers
XX CDS 33..2630
XX FT /tag= "a
XX FT /product= "protein kinase/protein phosphatase"
XX
XX PN WO200109316-A1.
XX
XX PD 08-FEB-2001.
XX
XX PF 28-JUL-2000; 2000MO-JP05061.
XX
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 18-OCT-1999; 99US-0159590.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 17-FEB-2000; 2000US-0183322.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX
XX PA (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;
XX Senoo C, Nezu J;

```

```

XX WPI: 2001-570286/64.
DR P-PSDB: AAG67610.
XX
XX New genes encoding proteins with protein kinase/protein phosphatase
XX activity, useful in the diagnosis and treatment of diseases -
XX
XX Claim 1: Page 67-76; 233pp; Japanese.
XX
XX The present sequence encodes a human protein kinase/protein phosphatase.
XX It is expected that the protein kinase/protein phosphatase gene
XX participates in signal transduction in cells. The protein kinase/protein
XX phosphatase polypeptides and polynucleotides are useful for developing
XX diagnostics and treatment agents for human and animal diseases. The
XX protein kinase/protein phosphatase polypeptides are useful as target
XX molecules in designing novel drugs. The protein kinase/protein
XX phosphatase polynucleotides are useful as a source of probes and
XX primers, which may be used to isolate homologous sequences.
XX
XX Sequence 2718 BP: 840 A; 586 C; 688 G; 604 T; 0 other;

alignment_scores:
    quality: 4409.00      length: 909
    ratio: 5.068          gaps: 2
    Percent Similarity: 95.710      Percent Identity: 94.829

alignment_block:
US-09-757-781-2 x AAH78369 ..

Align seg 1/1 to: AAH78369 from: 1 to: 2718

126 GIUValThrProSerValLeuArgAlaAsnMetProLeuHisValaArgar 142
3 GAGGTCACACCTTCAGTCTTCGAGCAAAATGCTCTTATGTTGACG 52
142 gSerSerAspProAlaLeu1leGIuSerThrSerValSerAspSera 159
53 CAGTATGTGACCCAGCTCTTAATTGGCTCTCCATCTGTGATGATGTA 102
159 snPheSerSerGIuGIuProSerArqLysAsnProThrArqTrpSerThr 175
103 ATTTTCTCTGTGAGAGGCTTCAGAAATAATCCACACGCTGTGTAACA 152
176 ThrAlaGIuPheLeuLysGIuAsnThrAlaGIuSerProLysThrCysAs 192
153 ACAGCTGCTCTCTCAAGCAGAACACTGCTGGAGATCTCTAAAGCCTGGA 202
192 pArgLysLysAspGIuAsnTyrArgSerLeuProArgAspThrSerAsnT 209
203 CAGGAGAAAGATGAAACTACAGAAAGCTCCCGCGGATACTACTACT 252
209 rpSerAsnGIuPheGIuArgAspAsnAlaArgSerSerLeuSerAlaSer 225
253 GGTCTTAACCAATTTCAGAGACAAATGCTCGCTGCTGTGAGTCCAGT 302
226 HisProMetValGIuLysTrpLeuGIuLysGIuGIuGIuGIuGIuGIuGIu 242
303 CACCAATGTGTGGGCAAGTGCAGAGAAACAGAACAGACAGATGAGATG 352
242 yThrGIuGIuAspAsnSerArqValGIuProValGIuHisAlaAspThrG 259
353 GACAGAAAGATTAACACTGCTGTGAACTGTGGACATGCTGACACCG 402
259 1LeuGIuHis1leProAsnPheSerLeuAspAspMetValLysLeuVal 275
403 GTTGGACATATACCCAACTTTCTGTGATGATGATGTAAGCTCGTA 452
276 GIUValProAsnAspGIuGIuProLeuGIu1leHisValaProPhe 292
453 GAAGTCCCAACAGATGAGGCGCTCTGGAAATCCATGATGAGCTTTACG 502
292 rAlaArgGIuGIuArgThrLeuGIuLeuValValLysArgLeuGIuLysG 309

```



```

|||||
503 TCGTCGAGGCGGCGAGAACCTGGGGTTATTAGTAAACGATTGGAGAAAG 552
309 IYGLIYLSAIAGLIUNHISGLIUNSLNLEUPHEARGLIUNASPCYSILEVAL 325
553 GTGTGAAGCTGACATGATAAATCTTTTCGTGAGAAATGATTGCATTGTC 602
326 ARGILEASNASPGLYASPLEUARGASNARGARGPHEGLIUNALAGLIHI 342
603 AGCATTTAAATGATGGCGACCTTCGAAATAGAAATTTGAACAAGCACAAACA 652
342 SMETHEARGGLINALAMETARGTHRPROILETTRPHEHISVALVALP 359
653 TATGTTTCGCCAAGCCATGCTACACCCATCATTTGGTTCCATGTGTTCC 702
359 FOALIALIASNLYSGLIUNTYRGLIUNLEUSERGLINSEGLIUNYASIN 375
703 CTGCAGCAAAATTAAGACAGTATGAACAACATCCCAAAGTGAAGAAAC 752
376 ASNTYRTYRSESRARGPHESESRPROASPSERGLINTYRILEASPSANR 392
753 AATTACTATTCAAGCCGTTTTCACCTTGACAGCCAGTATATTGACAAACAG 802
392 GSERVALIASNSERIALAGLYLEUHIESTHRVALGINARGALAPROARGLEUA 409
803 GAGTGTGAACAGTGCAGGGCTTCACACGGTGCAGAGACACCCGACTGA 852
409 SNHISPROPROGLIUNILEASPSERHISERARGLEUPROHISERIALA 425
853 ACCACCCGCTGACGACGATGACACTCTCAAGACTACCTCATAGCCGA 902
426 HISPROSERGLIYASPROBROSERALAPROALASERIALAPROGLIASNVA 442
903 CACCCCTCGGGAACACACCATCCGCTCCAGCCTCGGACATCAGAAATG 952
442 LPHESERTHRTHRVALSERSERGLITYRASNTHRLYSYLISLEGLYLYA 459
953 AATTAGTAGACTGTAAAGCATGGTTATTAACACCAAAAAAATAGCCMAA 1002
459 IGLIUNASNLIEGLINLEULYSGLIYTHRGLIUNLYLEUGLYPHESEITIE 475
1003 GCGTTNATTCACAGCTTAAGAAAGTACGAAAGGTTGGGATTCACACATC 1052
476 THRSEARGASPVALTHTRIEGLYLSERIALAPROILETYRVALLYSAS 492
1053 ACTTCCAGAGATGTAAACATAGTGGCTCAGCTCCAAATCTATGTGAAAA 1102
492 NILELEUPROARGGLYALALALALEGLINASPGIYARGLEULYSALAGLYA 509
1103 CATTTCTCCCGGGGGGGCGGCATTCAGATGCGGCATTAAGCGAGGAG 1152
509 SPARGLEULIEGLIUNVALIASNGLYVALASPLEUVALGLIYSSSERGLINU 525
1153 ACAGACTTATAGAGTAAATGAGATGATTAGTGGCAAAATCCCAAGAG 1202
526 GIUNVALVALSERLEULIUNARGSERTHRILYSMETGLUIGLYTHRVALSERIE 542
1203 GAAGTTGTTTCGCTGTGAGAAACACCAAGATGGAAGAACTGTGAGCCT 1252
542 UNLEUVALPHEARGGLINGLIUNASPLAPHEHISPROARGLIUNLEUASNLAG 559
1253 TCTGCTCTTTCGCCAGGAAGACGCTTCCACCCAAAGGAATGAATGACAG 1302
559 IUPROSERGLINMETGLINILEPROLYSGLIUTHRLYSALAGLIUNASPGIUNAS 575
1303 ACCCAAGCCAGATGCGAGATTCAAAAAAGAAACGAAACAGATGAGAT 1352
576 ILEVALLEUTHRPROASPGIYTHRARGLIUNPHELEUTHRPHIEGLIUNALP 592
1353 AATTGTTCTTACACCTGATGCGACACAGGAATTTCTGACATTTGAAGTCCC 1402
592 OLEUNASNASPSERGLYSERIALAGLYLEUCLYVALSERVALIYSGLYASNA 609
|||||

```

```

1403 ACTTACTGATTTCAGATCTGCAGGCGCTTGGTGTCACTGTCAAAAGTAAACC 1452
609 IYSERLYSGLIUNSNHISALIASPLEUGLIYILEPHEVALIYSSSERILEIE 625
1453 GGTCAAAAGAGAACACCGCAGATTGGGAATCTTTGTCAAGTGCATTATTT 1502
626 ASNGLIYALIALASERLYSASPGIYARGLEUARGVALIASNASPGLIUN 642
1503 AATGAGAGGACACCATCTAAAGATGGAAGCGCTTCGGGTGAATGATCAACT 1552
642 UILEALIASNGLIYGLUSERLEUENGLIYUSTHRASNGLINASPLAM 659
1553 GATAGCAGTAAATGAGAAATCCCTGTGGGCAAGACAAACCAAGATGCCA 1602
659 ETCGLIUTHRIUNARGARGSERMETSESRTHRGLIUNLYASNLYSARGIYMET 675
1603 TGGAAACCCTAAGACAGTCTATCTCTACGTAAAGCATTAACAGAGATG 1652
676 ILEGLIUNILEVALIALARGARGIIESERLYSCYSANGLIUNLEULYSE 692
1653 ATCCACCTTATTTGTCCAAAGAGAAATAGCAAGTCAATGACCTGCAAGTGC 1702
692 RPROGLYSESRPROPROGLIYPROGLIUNPROILEGLIUTHRIALIASPSA 709
1703 ACCTGGAGGCCCTCGAGACTGAGCTGCCATTTGAACAGCCTTGGATG 1752
709 SPARGLIUNARGARGIIESERHISERILEUITYRSEGLIYILEGLIUNLY 725
1753 ATAGAGAACCAAAATTTCCCATTCCTCTACAGTGGGATTTGAGGGCTT 1802
726 ASPGLUSERPROSESRARGASNALIALALEUSERARGIILEMETGLYIUNSE 742
1803 GATGAATCGGCCAGAGAAATGCTGCCCTCAGTAGATATG..... 1844
742 RGLIYSTYRGLINLEUSERPROTHRVALIASMETPROGLINASPSAPHRV 759
1845 .GGTAAATACACAGCTGTCCCTACAGTGAATGCCCCAAGATGACACTTG 1893
759 ALILEIIEGLIUNASPSARGLEUPROVALLEUPROPROHISLEUSERASP 775
1894 TCAITTTAGAGATGACAGGTTGCCAGTCTCTCTCACATCTCTGTGAC 1943
776 GLINSESRSESRSESRHISASPSVALIGLYPHEVALIUNALIASPAL 792
1944 CAGTCCCTCTCCAGCTCCCATGATGATGTTGGTTGTGACGCGACAGATGC 1993
792 AGLIYTHRTTPALIALYALALALIESERASPSERIALIASPCYSERIEUS 809
1994 TGGTACTTGGGCCAAGGCTGCAATCAGTGAATTCAGCCGACTGCTCTTTGA 2043
809 ERPROASPVALASPROVALLEUALAPHEGLINARGGLIYPHEGLIYARG 825
2044 CTCACAGATGTGATCCAGTTCCTTGCTTTTCAACGAGAAGATTTGGACGT 2093
826 GLINSERMETSESRGLIUNYSARGTHRILYSGLINPHESEASPLASERGLIUN 842
2094 CAGACT..... 2099
842 UNASPHEVALIYSESRTHRARGLYSSERLYSSERMETASPLEUGLIYILEALIA 859
2100 .....G 2100
859 SPGLIUTHRLYLSUNASNTHRVALIASPSGLINLYSALAGIYSEPROSER 875
2101 ACAGAGACTAAACTCATATACGTGATGACACCAAGAACAGGTTCTCCACG 2150
876 ARGASPVALIGLYPROSERLEUGLYLEULYSLYSSERSESRLEUGLIUNSE 892
2151 ACAGATGTGGGCTCTCCCGGGGTCTGAAGAAAGTCAAGCTCATTTGAGAG 2200
892 RLEUGLINTHRIALIALVALIAGLIUNVALIUNLEUASNGLYASPILEPROPH 909
2201 TCTGCAGACCGCAGTTGCCAGGCTGACTTTGAATGGGATATTCCTTTCC 2250

```

```

909 1sArgProArGProArG1le1leArgGlyArgGlyCysAsnGluSerPhe 925
2251 ATGCTCAGCGCGCGATATCATCAGAGCGAGGATGCAATGAGAGCTTC 2300
926 ATcAlAlAlAlAlAspLysSerTyrAspLysProAlaValAspAspAsp 942
2301 AGAGCTGCGCATCGCAAAATCTTATGATTAACCCGCGCTAGCATGATGATGA 2350
942 pgluglMetGluThrLeuGluGluAspThrGluGluSerSerArgSerc 959
2351 TGAAGGATGAGACCTTGCAAAAGACACAGCAAGAAAGTTCAAGATCAG 2400
959 lYArgGluSerValSerThrAlaSerAspGlnProSerHisSerLeuGlu 975
2401 GGAGAGAGTCTGTATCTCACAGCAGATGATCAGCCTCCACCTCTCGGAG 2450
976 ArgGluMetAsnGlyAsnGlnGluLysGlyAspLysThrAspArgGlySyl 992
2451 AGACAAATGAAATGGAACCAAGAAAGGATGATTAAGACTGATAGAAAAA 2500
992 sAspLysThrGlyLysGluLysLysLysAspArgAspLysGluLysAspL 1009
2501 GGATTAACCTGGAAGAAAGAAAGAAAGATAGATTAAGGACAGAGATTA 2550
1009 yMetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArg 1025
2551 AATGAAAGCCAAAGAGGAAATGCTGAGGCGCTTGAGGACATGTTCCAGC 2600
1026 PheGlyLysHisArgLysAspAspLys 1034
2601 CTTCGCAAACTGAGCCCGCAGAAAGAA 2627

seq_name: /SIDSL/9cgcdata/hold-9cgeneseg/9cgeneseg-emb1/NM2000.DAT: AAD01182
seq_documentation_block:
ID AAD01182 standard; cDNA; 2554 BP.
AC AAD01182;
XX
DT 04-OCT-2000 (first entry)
XX
DE PDZ domain-comprising PHIP (ephAin Interacting Protein) complex cDNA.
XX
KW PDZ domain; B class ephAin; Eph receptor tyrosine kinase; RTK; modulator;
cellular process; nerve cell interaction; regeneration of nerve cell;
axonogenesis; antiproliferative; proliferative disorder; treatment; GRIP;
differentiative disorder; PHIP; ephAin interacting protein; synAenin; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 306..2540
FT /tag- a
FT /product= "PHIP (ephAin Interacting Protein) complex"
FT /note= "Homologous to C. elegans PAR-3 protein"
XX
PN W0200031124-A2.
XX
PD 02-JUN-2000.
XX
PE 19-NOV-1999; 99WO-CA01101.
XX
PR 20-NOV-1998; 98US-0109158.
XX
PA (MOUN ) MOUNT SINAI HOSPITAL.
XX
PI Lin D, Pawson A;
XX
DR MPI: 2000-400038/34.
XX
DR P-PSDB; AATY1410.
XX
PT Isolated complex for treating proliferative or differentiative

```

```

PT disorders comprises B class ephAin and PDZ domain containing protein -
XX
PS Disclosure; Page 57-59; 59pp; English.
XX
CC The patent discloses a complex comprising of a B class ephAin and PDZ
CC domain containing protein. B class ephAin function as ligands for Eph
CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a
CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
CC binding site. This complex is used in methods to modulate the interaction
CC of a B class ephAin and PDZ domain containing protein and to identify
CC like, axonogenesis, nerve cell interactions and regeneration of nerve
CC cells. The complex is also useful for treating proliferative or
CC differentiative disorders associated with this protein complex.
CC The present sequence is the cDNA clone encoding PHIP (ephAin Interacting
CC Protein) complex, comprising three PDZ domains. This clone is isolated
CC by screening a lambdaEcl10.5 day mouse embryo expression library.
CC PHIP is closely related to Caenorhabditis elegans PAR-3 protein involved
CC in regulating polarity of the early embryo. PHIP cDNA fragment was used
CC as a probe, to isolate other transcripts like GRIP and synAenin proteins
CC from 10.5 day mouse embryo library.
XX
SO Sequence 2554 BP; 650 A; 697 C; 745 G; 462 T; 0 other;

```

## alignment\_scores:

Quality	Ratio	Length	Gaps
3527.00	4.892	743	0
Percent Similarity: 97.039	Percent Identity: 91.386		

## alignment\_block:

US-09-757-781-2 x AAD01182 ..

Align seg 1/1 to: AAD01182 from: 1 to: 2554

```

1 MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17
306 ATGAAAGTGACCGCTGTGCTCGGAGAGACCGGGTGCTGCGCGGG 355
17 YAspGlyHisMetLysValPheSerLeuLleGlnGlnAlaValThrArgT 34
356 AGATGGCCGATGAAGAGTTTTCAGCTTCATCCAGCAGCGGGTGACCGCT 405
34 YArgGlyAlaAlaLysAspProAsnTyrTrpLleGlnValHisArg 50
406 ACCGAAAGCGCGTGCCAGAGATCCAAACTCTGATACAGTCAATCCG 455
51 LeuGlnHisGlyAspGlyLleLeuAspLeuAspLleLeuCysAs 67
456 TTGGAGCATGAGATGAGGATTCCTAGACCTCGATGACATCCCTGTGA 505
67 PValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84
506 CGTTGCTGATGACAAAGACAGACTGTAGCATATTTGATGAAGAGATC 555
84 roHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
556 CCCACATGAGGAGATGTACAGCGCCAGCTCCAGGAGACCCAGAGT 605
101 ProGluLlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
606 CCAGAGATATTCGGCAGTGAAGTGGGACCAACAATGTTCTGCTTTTCA 655
117 nProTyrGlnAlaThrSerGluLleGluValAlaThrProSerValLeuArg 134
656 GCCTATCAAGCCCAAGCAAGTGAATTTGAGTACGCTTCAGTTCTTGGG 705
134 laAsnMetProLeuHisValArgSerSerAspProAlaLeuLleGly 150
706 CAATATAGCTCTTCATGTCGCGGAGGAGCAGCAGCTTAATCTGG 755
151 leuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167

```

756 CTTCCACTTCTGTCAgTgATTAACAACCTTTCTCAGAGAGAGCCCTCCAG 805  
167 glysAsnProthrArgTyrSerThrAlaGlyPheLeuLysGlnAsnT 184  
|||||  
806 GAAAAACCCACCGCGTGTCCAGCAGACAGTGGCTTTCTCAACAGACAGA 855  
184 hAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200  
856 CCCTGGAACTCCCAAAACCTGCAGACAGAAAGATGAAACTACAGA 905  
201 SerLeuProArgSprHisSerAsnTyrSerAsnGlnPheGlnArgAspAs 217  
906 AGCTTTCACAGGAGTCCACTACTGTCTCAACAGACTTCCAGCGAGCAAA 955  
217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTyrLeuc 234  
956 CGCCCGCTCTCCCTGAGCGCCAGCCCAATGGTATGACCGGTGGTGG 1005  
234 LuLysGlnGluAsnAspGluAspGlyTyrGluGluAspAsnSerArgVal 250  
1006 AGAAGCAGAAACAGATGAGAGAGGCACAGAAAGACACAGCCGAGTG 1055  
251 GluProValGlyHisAlaSprThrGlyLeuGluHisTleProAsnPheSe 267  
1056 GAGCCGGTTGGACATGCTGATACCGGATGGAGAACATGCCAACTTTC 1105  
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284  
1106 CCTCGATGATATGTAAAGCTCTACAAAGTCCCAACGATGGAGGGCCC 1155  
284 euGlyIleHisValValProPheSerAlaArgGlyArgThrLeuGly 300  
1156 TGGGAATCCATGTAGTGCCTTTCAGTCTCGAGCGGCAGAACATTCGGG 1205  
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisGluAsnTle 317  
1206 TTGTTAGTGAAGCGGTGGAGAAAGCGGTAAAGCTGAGCAAGAAACCT 1255  
317 upheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334  
1256 TTTTCATGAGAAATGACTGCTGTGAGAGATTAAAGATGGAGATCTTGAA 1305  
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350  
1306 ACAAGAGATTGGAGCAAGCAACATATGTTCCGCCAAGCTATCGCTGCG 1355  
351 ProIleIleThrPheHisValValProAlaAlaAsnLysGluGlnTyrG 367  
1356 CGGTCTATTGGTTCATGTGTGTCCTCGACAAACAGAGCAATATGA 1405  
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384  
1406 ACAACGTCCCAACGCGAAGAAACATCTACCTCCCGCGCGCTTCAGCC 1455  
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400  
1456 CTGACAGGCCACTGTGTGGCCAAACAGAGAGTGTGGCAACATGGCCCTCAA 1505  
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417  
1506 GCATGGCCAGACACCCACAGACTGATGACACCCAGACAGCGTGGATGC 1555  
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434  
1556 TCACCCCGCACTACCTCATAGTGTCTCAACGCTCAACCAACACACCCGAG 1605  
434 lApProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450  
1606 CCCCAGCTTGGCTCCACCCAGCTGTCTTAGTACCAACAGTGGCAGCTGTG 1655  
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467  
1656 TACAAACGAAAGAAATGAGCAAGGCTCAACATCCAGCTTAAGAAAGG 1705

467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484  
1706 TACAGAAAGACTGGGATTTCAAGCATCACCTCCCGGAGCTCACCATAGTGTG 1755  
484 LysSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500  
1756 GCTCAGCTCCCATTTATGTCAAGAAATATCTTCTCCAGAGGGGCTGCCATT 1805  
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517  
1806 CAGGATGGACACTCAAGCAGAGACCGGCTAAATAGAGTCAAATGAGT 1855  
517 lAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerT 534  
1856 AGATTTAGCAGGCAAAATCCAGAGAGAACTGTTTCCCTGTGGAGAACGA 1905  
534 hrlYsMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550  
1906 CCAAGATGGAGGGAGCTGTGAGCTTCTGTGTCTTCAGAGAAAGAGCT 1955  
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProly 567  
1956 TTCCACCCCAAGGGAATGATGCTGAACCAACAGCATGCAGACTCCAAA 2005  
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584  
2006 AGAAACGAAGCTGAGATGAGAGCTGTTCACACCCGATGGTATACA 2055  
584 rglLupheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600  
2056 GGGAGTTTCTCATTGTGAAGTTCACCTGATGATGACTCGAGATTCGACAG 2105  
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617  
2106 CTTGTGTCACTGTCAAGGGAGACCGTTCCAAAGAAACCCAGATTT 2155  
617 uGlyIlePheValLysSerIleIleAsnGlyAlaAlaSerLysAspG 634  
2156 GGGGATCTTCTGTTAAATCCATTATCAATGGTGGAGCTCATCTAAAGATG 2205  
634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650  
2206 GAAAGCTGAGGGTAAATGACCACTGATAGCTGTGAAGGAGAAATCTCTTA 2255  
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667  
2256 CTGGGCAAAAGCCAAACAGAAAGCCATGAGACTTACAGAGGTCCATGTC 2305  
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684  
2306 CACGAGAGGCAACAGCGTGCATGATTCACGCTCATTTGTGGAGGCGGA 2355  
684 lSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700  
2356 TCACAGCAATGTAAAGAGCTGGGTCTCTGTGGAGCCCTGCTCACCTTAG 2405  
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717  
2406 CTGGCCATCCGACAGAACTGGATGACCGAGAAACGAGAGATCTCACACTC 2455  
717 rLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAlaA 734  
2456 CCTCTACAGTGGATTCGAGGGGCTGAGATGAGTGGCCCAACGAGAAACGAG 2505  
734 lAlaLeuSerArgIleMetGlyLysSerGly 743  
2506 CACTGAGCAGATATATGGTGAATCAGGA 2534  
seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-embL/NA2001A.DAT:AAH72793  
seq\_documentation\_block:  
ID AAH72793 standard; cDNA; 1800 BP.



|||||  
1095 TGCAGCAATTCGATTCCTTAAACACCAAAATCAAAAGCATGATT 1144  
855 eucllytlealaaspgluThrlyleuasnThrValaaspGlnlyala 871  
1145 TAGCTATAGCTGACGACGACTAACCTCAATACAGTGCACCGAAAGCA 1194  
872 GlySerProSerArgAspValGlyProSerleuGlyleuLysSerSe 888  
1195 GGTTCTCCACGACGAGATGGGGCTCTCCCTGGGTCTGAAGAAGTCAAG 1244  
888 rSerleuGluSerleuGlnThrValaAlaGluValThrleuasnGly 905  
1245 CTGCTGGAGAGCTCTGCAGACCCGAGTTCGCCGAGTGACTTGAATGGG 1294  
905 spIleProPheH1sArgProArgProArgIleIleArgGlyArgGlyCy 921  
1295 ATATTCTCTTCATCGTCACGCGCGCGGATTAATCAGAGCGAGGGGATGC 1344  
922 AsnGluSerPheArgAlaAlaIleAspLysSerTyrAspLysProAla 938  
1345 AATGAGAGCTTACAGAGCTGCCATCGACAATCTTATGATAACCCCGGT 1394  
938 IAspAspAspAspGluGlyMetGluThrleuGluGluAspThrGluGlu 955  
1395 AGATGATGATGATGAAGCATGAGACCTTGAAAGAACACAGAACAA 1444  
955 eSerArgSerGlyArgGluSerValSerThrAlaSerAspGlnProSer 971  
1445 GTTCAAGATCAGGAGAGACTGTGTATCCACAGCCAGTGCACGCTTCC 1494  
972 HisSerleuGluArgGlnMetAsnGlyAsnGlnGlyLysAspLysTh 988  
1495 CACTCTCTGAGAGACAAATGAATGGAACCAAGAAAGATGATTAGAC 1544  
988 rAspArgLysLysAspLysThrGlyLysGlyLysLysLysAspArgAspL 1005  
1545 TGATAGAAAAAGATAAACTGAAAAAGAAAGAAAGATAGAGATA 1594  
1005 yAsGluLysAspLysMetLysAlaLysLysGlyMetLeuLysGlyLeuGly 1021  
1595 AGGAGAAAGATTAATAATGAACCAAGAGAGGATGCTGAAGCGCTTGGA 1644  
1022 AspMetPheArgPheGlyLysHisArgLysAspAspLys 1034  
1645 GACATGTTCAGCTGCCAAACTGAAGCCGAGAGAGAGA 1683  
seq\_name: /SID1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AA527320  
seq\_documentation\_block:  
ID AA527320 standard; cDNA; 823 BP.  
XX  
AC AA527320;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE cDNA encoding novel signal transduction pathway protein, Seq ID 355.  
XX  
KM Neuropeptide; cytosolic; dermatological; immunosuppressive; tumour;  
KM antiinflammatory; anti-HIV; antibacterial; antinflammatory; cancer;  
KM immune system disorder; rheumatoid arthritis; inflammatory condition;  
KM organ transplant rejection; infection; hepatitis C; blood disorder;  
KM sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KM neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KM chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KM cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KM reproductive system; gastrointestinal; liver disorder; AIDS; ss;  
KM acquired immune deficiency syndrome.  
XX  
OS Homo sapiens.  
XX  
PN WO200154733-A1.  
XX

PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US01312.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0198874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0203515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234598.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.



393 erValaasneralaglyleuHsthrValglalargalaproargleuasn 409  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
406 GTGTGAACAGTCAGGGCTTCACACGTCGACAGAGACCCCGACGTGAAC 455  
CC the proteins are collectively known as colon cancer antigens. The colon  
410 HisProProgluglnIleasPserHisSerArgLeuProHisSerAlaH 426  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
CC cancer antigens have cyrostatic activity and can be used in gene  
456 CACCCGCTGAGCAGATGACTCTCACTCAAGACTACCTCATAGCCGACA 505  
CC therapy and vaccine production. N and P may be used in the prevention,  
426 sProSerGlyLysProProSerAlaProAlaSerAlaProGlnAsnValP 443  
CC diagnosis and treatment of diseases associated with inappropriate P  
506 CCCCCTGGGAAAAACACATCCGCTCCAGCTCGGCACCTCAAAATGTAT 555  
CC expression. For example, N and P may be used to treat disorders  
443 heserThrThrValSerSerGlyTyrAsnThrLysLysIleGlyLysArg 459  
CC associated with decreased expression by rectifying mutations or deletions  
556 TTAGTACGACTGTACGACAGTGTATATACACAAAAAATAGCAAGAG 605  
CC in a patient's genome that affect the activity of P by expressing  
460 leuAsnIleGlnleuLysGlyThrGlnGlyLeuGlyLysSerIleTh 476  
CC inactive proteins or to supplement the patients own production of P.  
606 CTTAATATCCAGCTTAAGAAAGGTACGAAGGTTGGGATTCAGCATCAC 655  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
476 rSerArgAspValThrIleGlyLysSerAlaProIleTyrValLysAsn 493  
CC by inserting the nucleic acids into a host cell and culturing the cell  
656 TTCCAGAGATGATACATAGTGGCTCAGCTCCATCTATGTGAAAAACA 705  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
493 leuLeuProArgGlyAlaAlaIleGlnAspGlyArgLeuLysAlaGlyAsp 509  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
706 TTCTCCCGCGGGGGCGGCATTCAGAGATGCGCAGCTTAAGCAGAGAC 755  
CC present invention.  
510 ArgLeuIleGlnValaAsnGlyValaAspLeuValGlyLysSerGlnGlu 526  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
756 AGACTTATAGAGGTAAATGAGTTAGTTAGTGCGCAAAATCCCAAGAGA 805  
CC missing at time of publication, meaning no sequences are present for  
526 uValValSerLeuLeu 531  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
806 AGTGTGTCGCTGTG 821  
XX  
seq\_name: /SIDSL/gcgcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH33594  
seq\_documentation\_block:  
ID AAH33594 standard; cDNA; 823 BP.  
XX  
AC AAH33594:  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE Human colon cancer antigen encoding cDNA SEQ ID NO:650.  
XX  
KM Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KM colorectal carcinoma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200122920-A2.  
XX  
PD 05-APR-2001.  
XX  
PE 28-SEP-2000; 2000WO-US26524.  
XX  
PR 29-SEP-1999; 99US-0157137.  
PR 03-NOV-1999; 99US-0163280.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Barash SC, Blise CE, Rosen CA;  
XX  
DR WPT: 2001-233537/24.  
XX  
P-PSDB: AAG74163.  
XX  
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX  
PS useful for preventing, diagnosing and/or treating colorectal cancers -  
XX  
Claim 1; Page 2675; 9803pp; English.

XX  
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cyrostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
SO Sequence 823 BP; 249 A; 183 C; 201 G; 188 T; 2 other;  
XX  
alignment\_scores:  
Quality: 1397.00 Length: 272  
Ratio: 5.174 Gaps: 0  
Percent Similarity: 99.265 Percent Identity: 99.265  
XX  
alignment\_block:  
US-09-757-781-2 x AAH33594 ..  
XX  
Align seg 1/1 to: AAH33594 from: 1 to: 823  
XX  
260 leuGlnHstIleProAsnPheserLeuAspAspmetVallyLeuValGI 276  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
6 TTGGAGCATATACCCCACTTCTCTGTGATGATATGGTAAAGCTCGTGA 55  
CC  
276 uValProAsnAspGlyLysProleuGlyIleHisValValProPhesera 293  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
56 AGTCCCAAGATGAGAGGGCCTCGGGAATCCATGTAGTGGCTTTCAGTG 105  
CC  
293 laarGlyGlyArgThrLeuGlyLeuLeuValLysArgLeuGlyLysGly 309  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
106 CTCGAGGCGGCGAGAACCCCTGGGTTATATGATTAACCATTTGCAGAAAGCT 155  
CC  
310 GlyLysAlaGlnHstGluAsnLeuPhearGlnAsnAspCysIleValAr 326  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
156 GTTAAAGCTGAACTGAAATCTTTTCGTGAGATATATTCATTTGCTGAC 205  
CC  
326 gIleAsnAspGlyAspLeuArgAsnArgArgPhaGlnGlnAlaGlnHstM 343  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
206 GATTAAATGATGGCAGCTTCGAAATAGAGATTTGAAACACACACACTA 255  
CC  
343 elPhearGlnAlaMetArgThrProIleIleTrrPheHisValValPro 359  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
256 TGTTCGCCAAGCCATGCGTACACCATTCATTTGTTCCATGTGTTCCCT 305  
CC  
360 AlaAlaAsnLysGlnGlnGlnGlnLeuSerGlnSerGlnLysAsnAs 376  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
306 GCAGCAATATAAAGAGCAGTATGAACACTATCCCAAAAGTGGAAGAACAA 355  
CC  
376 nTyTyrSerSerArgPheserProAspSerGlyTyrIleAspAsnArgS 393  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
356 TTACTATTTCAAGCCGTTTTCAGCTTCAGCAGCAGCATATATTACAAACAGA 405  
CC  
393 erValaasneralaglyleuHsthrValglalargalaproargleuasn 409  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
406 GTGTGAACAGTCAGGGCTTCACACGTCGACAGAGACCCCGACGTGAAC 455  
CC  
410 HisProProgluglnIleasPserHisSerArgLeuProHisSerAlaH 426  
CC ||||||||||||||||||||||||||||||||||||||||||||||||  
456 CACCCGCTGAGCAGATGACTCTCACTCAAGACTACCTCATAGCCGACA 505  
CC

CC can be used as targets in developing anti-cancer agents such as  
 CC chemotherapeutic compounds. The markers can also be used as targets in  
 CC developing treatments for cancer, particularly those cancers which  
 CC display resistance to agents and exhibit expression of the markers. The  
 CC anticancer agents developed by the novel method can be used to treat  
 CC cancer. Probes based on the markers can be used to detect transcripts or  
 CC genomic sequences corresponding to the markers, in the identification of  
 CC cells or tissues which mis-express the protein. Cancers which may  
 CC be targeted include carcinoma (e.g. squamous cell carcinoma),  
 CC sarcoma (e.g. fibrosarcoma), leukaemia (e.g. lymphocytic leukaemia),  
 CC lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's disease and  
 CC tumours (e.g. glioma). The present sequence is one of the 1046  
 CC novel cancer cell markers.

XX  
 SQ Sequence 2100 BP; 649 A; 460 C; 544 G; 439 T; 8 other;

alignment\_scores:  
 Quality: 1187.50 Length: 849  
 Ratio: 2.385 Gaps: 25  
 Percent Similarity: 58.657 Percent Identity: 36.867

alignment\_block:  
 US-09-757-781-2 x AAS60861 ..

Align seg 1/1 to: AAS60861 from: 1 to: 2100

```
seq_name: /SIDSL/gcgdata/hold-geneseg/geneseg-n-emb1/NA2001A.DAT.AAS60861
seq_documentation_block:
ID   AAS60861 standard; cDNA; 2100 BP.
XX
AC   AAS60861;
XX
DE   29-JAN-2002 (first entry)
XX
DE   Human cancer agent-resistance marker #520.
XX
KW   Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
KW   squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
KW   lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
KW   Hodgkin's disease; glioma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200179556-A2.
XX
PD   25-OCT-2001.
XX
PF   13-APR-2001; 2001WO-US12132.
XX
PR   14-APR-2000; 2000US-197538P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J, Brown JL, Bolt A, Van Huiffel C;
XX
DR   WPI; 2001-602933/68.
XX
PT   Novel nucleic acid, used as a marker to determine the effectiveness of
XX   using TAXOL to treat cancer cell growth in individuals -
PS   Claim 1; Page 370; 527pp; English.
XX
CC   The invention relates to 1046 novel nucleic acids which are used as
CC   markers for determining the sensitivity of a cancer cell to the
CC   anticancer agent TAXOL. Cancer cells can be treated with TAXOL when
CC   they are shown to express one of the 242 sensitivity markers or the
CC   cells are shown not to express one of the 804 resistance markers.
CC   The methods can be used to determine the effectiveness of TAXOL
CC   in the treatment of cancer cell growth in an individual. The markers
```

```
223 SerAlaSerHisPromeIValGlyIleTrpLeuGluysGingInAs 239
|||||
49 AGCGCTTCACACCTCGG.....TGCTTAGAGTGTGAACCGGATG 88
|||||
239 pGluAspGlyThrGluGluAspAsnSerArgValGluProValGlyHisA 256
|||
89 TTCGAGGCTACGACGAGAACTTGGAAGACAGA...GAACTTTTGATGGTGT 135
|||||
256 lAspThrGlyLeuGluHisIleProAsnPro.....SerLeuAspAsp 270
|||||
136 TCACAGACAGAACTACTAACTTCGCCAAGACATAGAGACATGTGATAT 185
|||||
271 MetValIysLeuValGluValProAsnAspGlyIleProLeuGlyIleH 287
|||||
186 ARGACAGAGAACAGTGGAGATTTCTGGGAAGAGGCCCATTTGGCAATACA 235
|||||
287 sValValPro...PheSerAlaArgGlyIleArgThrLeuGlyLeuLeu 303
|||||
236 TGTAGTGCCTTTCTTTCTGATCTGTGAGTGAAGATCTTAGAGACTCTCA 285
|||||
303 AlIysArgLeuGluIysGlyIysAlaGluHisGluAsnLeuPheArg 319
|||||
286 TCAGAGGCAATTGACAGACACAGAGTCCAGCGGAGGACTATTTCAC 335
|||||
320 GluAsnAspCysIleValArgIleAsnAspGlyAspLeuArgAsnArgAr 336
|||||
336 GAAATGCAATGTATGTGTAATAAATCAACAATGTGATCTGTGACACAAAC 385
|||||
336 grHeGluGluAlaGluHisIleMerPheArgIleAlaMetArgThrProIle 353
|||||
386 CTTTGCTCAGGCTCAAGATGTCTTCGCCAGGCAATGAATATCCCAAGTG 435
|||||
353 lEtrPheHisValValProAlaAlaAsnLysGluGlnIleTrpGluGlnLeu 369
|||
436 TGCTTCGTCACGCTCAAGATGTCTTCGCCAGGCAATGAATATCCCAAGTG 482
|||
370 SerGlnSerGluLysAsnAsnIleTrpSerSerArgPheSerProAspse 386
|||
483 TCAGTCAATGGCTCTTAACATTTTGTGAATGAATGATGGCTTTTGAA 532
|||
386 rGlnIleAspAsnArgSerValAsnSerAlaGlyLeuHisThrValG 403
|||
533 AACCAAGTGCCTTCCTTCATGTAAGAAATGGGACTAAACACA.... 578
|||
403 lnaGAlaProArgLeuAsnHisProGluGlnIleAspSerHisSer 419
|||
```





```

996 GtlyvsgtlytlysAspArgAspIysgltlyAsplysmetlysal 1012
      |||||.....::: .....::: ... |||
1953 GTCAAGCAGGAAAAAAGCCAAAGACAGAACTCAATCCAAGAAGAAAT 2002
1012 atytsAglymettleutlysGlyLeuGlyAspMetPheArgPheGlyLysH 1029
      ||||| ..|||.....||| ..|||.....|||
2003 AAGAGAG.....AAGGCCTTCGGCGCATCTGAGATTGTGGAAGA 2043
1029 ISArtylsasp.....AsplyllecgtlystrhglyLyslleYs 1042
      ::::::||| ..|||.....||| ..|||.....|||
2044 AGAAAGAGATPAAGGTGCAAGAGCTGACGAGAAAGTACTCTGAAA 2090
seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AS60956
seq_documentation_block:
ID_AAS60956 standard; cDNA; 2100 BP.
XX
AC_AAS60956;
XX
DT_29-JAN-2002 (first entry)
XX
DE_Human cancer agent-resistance marker #615.
XX
KW_Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
KW_squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
KW_lymphocytic leukaemia; lymphoma; plasmacytoma; reticulum cell sarcoma;
KW_Hodgkin's disease; glioma; ss.
XX
OS_Homo sapiens.
XX
PN_WO200179556-A2.
XX
PD_25-OCT-2001.
XX
PF_13-APR-2001; 2001WO-US12132.
XX
PR_14-APR-2000; 2000US-197538P.
PA_(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI_Lillie J, Brown JL, Bolt A, Van Huffel C;
PS_WPI: 2001-602993/68.
XX
PT_Novel nucleic acid, used as a marker to determine the effectiveness of
XX_treating TAXOL to treat cancer cell growth in individuals -
XX
Claim 1; Page 458-459; 527pp; English.
XX
XX The invention relates to 1046 novel nucleic acids which are used as
XX_markers for determining the sensitivity of a cancer cell to the
XX_antitancer agent TAXOL. Cancer cells can be treated with TAXOL when
XX_they are shown to express one of the 242 sensitivity markers or the
XX_cells are shown not to express one of the 804 resistance markers.
XX_The methods can be used to determine the effectiveness of TAXOL
XX_in the treatment of cancer cell growth in an individual. The markers
XX_can be used as targets in developing anti-cancer agents such as
XX_chemotherapeutic compounds. The markers can also be used as targets in
XX_developing treatments for cancer, particularly those cancers which
XX_display resistance to agents and exhibit expression of the markers. The
XX_antitancer agents developed by the novel method can be used to treat
XX_cancer. Probes based on the markers can be used to detect transcripts or
XX_genomic sequences corresponding to the markers, in the identification of
XX_cells or tissues which mis-express the protein. Cancers which may
XX_be targeted include carcinoma (e.g. squamous cell carcinoma),
XX_sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic leukaemia),
XX_lymphoma, plasmacytoma, reticulum cell sarcoma, Hodgkin's disease and
XX_tumours (e.g. glioma). The present sequence is one of the 1046
XX_novel cancer cell markers.
XX
Sequence 2100 BP; 649 A; 460 C; 544 G; 439 T; 8 other;

```

alignment_scores:	
Quality: 1187.50	Length: 849
Ratio: 2.385	Gaps: 25
Percent Similarity: 58.657	Percent Identity: 36.867
alignment_block:	
US-09-757-781-2 x AAS60956	..

Align seg 1/1 to: AAS60956 from: 1 to: 2100

223 SerAlaSerHisPrometValGlyLysThrPheLysGlnGlnGlnAs 239  
49 AGCCCTTCACACCCCTGG.....TGGTAGAGTCTGAAACCGGATG 88  
239 pGluAspGlyThrGlnGluAsnAspSerArgValLupProValGlyHisA 256  
89 TTCACGGGTCAACGCAAGACTTGGAGACAGA..GAAGTTTGANTGGTG 135  
256 laaSPthrGlyLeuGlnHisIleProasnPhe.....SerLeuaspSP 270  
136 TACAGACAGAGCACTACTACTTCGCGCAAGAACATAGACACATTGAGTGAT 185  
271 MetValLysLeuValGluValProaspnsprGlyProLeuGlyLysI 287  
186 ATGCAACAAACAGTGGATATTTCTGGGGAAGAGGCCCATTTGGGATACA 235  
287 sValValPro...PheSerAlaArgGlyGlyArgThrLeuGlyLeuVal 303  
236 TGTAGTGGCCCTTCTTTTATCTCTGACGTGGAGAAATGTCTAGGACCTTCA 285  
303 aLysArgLeuGlnLysGlyGlyLysAlaGlnHisGluAsnLeuPheArg 319  
286 TCCGAGGCAATTGAACACAAACAGCAGGCTCCAAACCGGAGGAGCAATTTCAC 335  
320 GluAsnAspCysIleValArgIleAsnAspGlyAspLeuArgAsnArg 336  
336 GAAATATGAATATATTGTAAATAATCAACAATGTGGATCTGTAGCAAAAC 385  
336 pGheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThrProIle 353  
386 CTTTGCTCAGGCTCAAGATGCTTCCCGCCAGGCAATGAATCTCCAAAGT 435  
353 leTPRPheHisValValProAlaAlaAsnLysGlnGlnIleArgGlnGln 369  
436 TGCCTCTCAGTGGCTGCTCTCCACAAACCGTGAAACAGTATGATAAAG... 482  
370 SerGlnSerGluLysAsnAsnArgLysSerSerArgPheSerProasp 386  
483 TCAGTCATTTGGCTCTTTAACATTTTGTATTAATGATGGCCTTTTGA 532  
386 rGlnIleTrIleAspAsnArgSerValAsnSerAlaGlyLeuHisThrValG 403  
533 AACCAAAATGGCGCCTCCGTGCTCCATGGAAATACGGGACATAAGACA.... 578  
403 lnaArgAlaProArgLeuAsnHisProProGlnGlnIleAspSerHisSer 419  
579 .....GCAAACTCTACAGGAACCGATAGTCTGAAACACAGATGCATGACGT 623  
420 ArgLeuProHisSerAlaHisProser.....GlyLysProP 432  
624 TCCCTGCACAAACAAAGAGTCCCGGATACCAAGGCTGGAGGAAAC 673  
432 oSerAlaProAlaSerAlaProGlnAsnValPheSerThrValSerS 449  
700 TGGGATTTGGCGACGACATAAAATGCAAAAGAAATTAAGATTGACGTAAAG 749  
466 LysGlyThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrI 482  
150 AAAGCCCTGGAAGACTGGTTTCACTGTGGTTACACAGAGACTCTTCAT 799

[illegible]

AC - AAS60963;

XX 29-JAN-2002 (first entry)

XX Human cancer agent-resistance marker #622.

XX Human: cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;  
 KW squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;  
 KW lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;  
 KW Hodgkin's disease; glioma; ss.

OS Homo sapiens.

PN WO200179556-A2.

PD 25-OCT-2001.

PF 13-APR-2001; 2001WO-US12132.

PR 14-APR-2000; 2000US-197538P.

PX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Lillie J, Brown JL, Bolt A, Van Huffel C;

DR WPL: 2001-602933/68.

XX Novel nucleic acid, used as a marker to determine the effectiveness of  
 PT using TAXOL to treat cancer cell growth in individuals -

PS Claim 1; Page 464-465; 527pp; English.

CC The invention relates to 1046 novel nucleic acids which are used as  
 CC markers for determining the sensitivity of a cancer cell to the  
 CC anticancer agent TAXOL. Cancer cells can be treated with TAXOL when  
 CC they are shown to express one of the 242 sensitivity markers or the  
 CC cells are shown not to express one of the 804 resistance markers. The  
 CC methods can be used to determine the effectiveness of TAXOL.  
 CC In the treatment of cancer cell growth in an individual. The markers  
 CC can be used as targets in developing anti-cancer agents such as  
 CC chemotherapeutic compounds. The markers can also be used as targets in  
 CC developing treatments for cancer, particularly those cancers which  
 CC anticancer agents developed by the novel method can be used to treat  
 CC cancer. Probes based on the markers can be used to detect transcripts or  
 CC genomic sequences corresponding to the markers. In the identification of  
 CC cells or tissues which mis-express the protein. Cancers which may  
 CC be targeted include carcinoma (e.g. squamous cell carcinoma),  
 CC sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic leukaemia),  
 CC lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's disease and  
 CC tumours (e.g. glioma). The present sequence is one of the 1046  
 CC novel cancer cell markers.

SO Sequence 2100 BP; 649 A; 460 C; 544 G; 439 T; 8 other;

alignment\_scores:

Quality: 1187.50

Ratio: 2.385

Percent Similarity: 58.657

Length: 849

Gaps: 25

Percent Identity: 36.867

alignment\_block:

US-09-757-781-2 x AAS60963 ..

Align seg 1/1 to: AAS60963 from: 1 to: 2100

223 SerAlaSerHisProMetValGlyLysTrpLeuGlyLysGlnGlnGlnAs 239

49 AGCGGTTACACCTCG.....TGGTTAGAGTCTGAAACCGGATG 88

239 PGLuAspGlyThrGlnLysPAsnSerArgValGluProValGlyLHisA 256

89 TTCAGAGGTCACGACGAACCTTGGAAACAGACA...GAAGTTTGAATGTTG 135

256 laAsPThrGlyLeuGlnHisIleProAsnPhe.....SerLeuAspAsp 270

136 TACAGACAGAACTACTAATCTCCGACAGAACTAAGACACATTTGAGCTGAT 185

271 MetValLysLeuValGluValProAsnAspGlyGlyProLeuGlyLLeuI 287

186 ATGACAAAGAACAGTGGAGATTCTGGGGAAGGAGGCCCATTTGGAAATACA 235

287 SValValPro...PheSerAlaArgGlyGlyArgThrTrpLeuGlyLeuVal 303

236 TGTAGTCCCTCTTTTCATCTGTGATGGAAGAGATTCTAGAGCTCTTCA 285

303 aLLysArgLeuGlyLysGlyLysAlaGlnHisGlnAsnLeuPheArg 319

286 TCCGAGGCAATTGAAACACAGCAGGTCACACCGGAGGAGCATATTTCAC 335

320 GLuAsnAspCysIleValArgIleAsnAspGlyAspLeuAlaGlnAsnArg 336

336 GAAATGAATGATTGTGTAATAATCAACATGTGGATCTGTAGACAAAC 385

336 pPheGlnAlaGlnHisMetPheArgGlnAlaMetArgThrProIleI 353

386 CTTCGCTAGGCTCAAGATGCTTCGCGACAGCAATGAATCTCCAGTG 435

353 laTrpPheHisValValProAlaAlaAsnLysGlnGlyTrpGlnGlnLeu 369

436 TCCTCTCCACGTCCTCTCCACAAAACCTGAAACGTATGAAAG... 482

370 SerGlnSerGlyLysAsnAsnTrpTyrSerSerArgPheSerProAsp 386

483 TCAGTCATGGCTCTCTTAACATTTTGGTAATAATGATGGCTTTGAA 532

386 rGlnTrpIleAsnAsnArgSerValAsnSerAlaGlyLysHisThrValG 403

533 AACCAAGTGGCGCTCTCCATGGAATGGAGCTGAAGACA... 578

403 LnaArgAlaProAlaGlyLeuAsnHisProGlnIleAspSerHisSer 419

579 ....GCAAAATTCACAGAAACGATAGTCGTGAACACATGCATCAGCT 623

420 ArgLeuProHisSerAlaHisProSer.....GlyLysProPr 432

624 TCCTGCAACAAACAAAGAGTCCGAGATACCAAGCTGGAGGAAACCC 673

432 oSerAlaProAlaSerAlaProGlnAsnValPheSerThrThrValSer 449

674 ATCTCTCCCTCACTCTCGCT.....CTCA 699

449 erGlyTyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLys 465

700 TGGGATTTGGCAGCAATAAAATGCAAAAGAAATTAAGATTGACCTTAAG 749

466 LysGlyThrGlnGlyLeuGlyPheSerIleThrSerArgAspValThrI 482

750 AAAGCCCTGAAAGACTGTGTTCACTGTGTACAGAGACCTTCCTCAT 799

482 eGlyLysSerAlaProIleTrpValLysAsnIleLeuProArgGlyAlaA 499

800 ACATGTCCTCCGTCCTTTTGTAAACATTTTACCAAGAGGAGCAG 849

499 laIleGlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsn 515

850 CTATATTAAGATGGCGCTCAATTCAGGAGCAGAAATTTTGGAGCTTAAT 899

516 GlValAlaSerLeuValGlyLysSerGlnGlnGlnValValSerLeuLeuAr 532

900 GGGAGAGATGTCAACCGGACCAACCGAGAAAGCTTGGCCATCTCAG 949

532 gSerThrLysMetGlnGlyThrValSerLeuLeuValPheArgGlnGlnA 549

950 GAGCACCAAGGAGGAGAGACAGCATCGCTGTCAFTGCCGCCCAAGAG 999





```

1022 ..... 1022
632 SAAPGIYARGLuEArGVAlAsnAspGlnLeuIleAlaValAsnGlyGluS 649
||||| .....
1023 .GATGTCGCTCGAATGATGATGACCAAGCTGATTGCAATTATGGGAAAT 1071
649 erLeuEngLyLysThrAsnGlnAspAlaMetGluThrLeuArGArGSer 665
||||| .....
1072 CTCTTTGGGAAGTCCAAACCAAGCAAGCTATGGAACACTTACGGCGTCA 1121
666 MetSerThrGluGlyAsnLysArGlyMetIleGlnLeuIleValAlaAr 682
||||| .....
1122 ATGTCATGAGGGAACATCGAGGATGATCAGTGGTATTCGAG 1171
682 gaRGISeRlycSaSnGluLeuLysSerProGlySerProGlyP 699
||||| .....
1172 GAGC.....C 1176
699 roGluLeuProIleGluThrAlaLeuAspAraGluArGArGLeSer 715
||||| .....
1177 CAGAGAGACCAATG..... 1190
716 HisSerLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArGAs 732
||||| .....
1191 .....GAGATCTCTGACAGGTG 1207
732 nAlaAlaLeuSerArGIleMetGlyGluSerGlyLysTyrGlnLeuSerP 749
||||| .....
1208 TGGGGCATTTTCCAGCCATGCTTGAGAAC.....TGTCMAAATGCTG 1251
749 roThrValAsnMetProGlnAspAspThrValIleIleGluAspAspArG 765
||||| .....
1252 TAACCACTCTAGCGAATGATATAGTATC..... 1283
766 LeuProValLeuProProHisLeuSerAspGlnSerSerSerSerHi 782
||||| .....
1284 .....CTGCATCCACTTGGCCTTGCAGTCCACA 1312
782 SAAP...AspValGlyPheValThrAlaAspAlaGlyThrTyrAlaLysA 798
||||| .....
1313 AGCAAAACAGAAAGGTCTATTGCTGCCCAATGACGA.....TGGGCCGAGA 1359
798 laAlaIleSerAspSer..... 803
||| .....
1360 GTGAAGTTCCACCTCTCCAAACACACATTCGCTCGGATGGGCCTC 1409
804 AlaAspCysSerLeuSerProAspValAspProValLeuAlaPheGlnAr 820
||||| .....
1410 GAAGATTACAGCCACAGCTCTGGGGTGATTACAGCACTATATTTTCCAGA 1459
820 gGIuGlyPheGlyArGlnSerMetSerGlyLysArGThrLysGlnPheS 837
||||| .....
1460 TCAGCACATCACTTCAGATCTGTGACACGGCCAG..... 1496
837 eRaSpAlaSerGlnLeuAspPheValLysThrArGlySerLysSerMet 853
||||| .....
1497 .....CAGCTGATCAATTAATTGAAGCCCTCGAAGACGATG 1535
854 AspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAspAspGlnLy 870
||||| .....
1536 GACCTT...GTGCCAGATGAAACAGCAAGTTCATTCATTTGCTGACAAAA 1582
870 saLaGlySerProSerArGAspValGlyProSerLeuGlyLeuLysLysS 887
||||| .....
1583 ATCGGAATCTCCAAAGAAATTTTGTCCAACTCTGGTTTGAAGAAAGT 1632
887 eRSeRLeuEngLysSerLeuGlnThrAlaValAlaGluValThrLeuAsn 903
||||| .....
1633 CCAGCTCTCTGGAGAGCTGACAGCTGACGCGGAGGAGAGAGAAT 1682
904 GLYAspIleProPheHisArGProArGProArGIleIleArGGLyArGAl 920
||||| .....

```

```

1683 ...GACCTTCTTTTCACAGGCCCGCCGACATGTTGAGCGCAGG 1729
920 YcYAsnGluSerPheArGAlaAlaIleAspLysSerTyrAspLysProA 937
||||| .....
1730 CTGCAATGACAGCTTTGACAGCCATTCGACAAATCTTACATGAGACT. 1778
937 laValAspAspAspAspGluGlyMetGluThrLeuEngLysAsp..... 951
||||| .....
1779 .....GAGCAATATAGAAAGCTGACGGrCTG 1802
952 ThrGluGluSerSerArGSerGlyArGluSerValSerThrAlaSerAs 968
||||| .....
1803 TCTGATAGAGCTTCTCACTCTGGCCAAAGAGCTCTGAATGTGAGTCTGC 1852
968 pGlnProSerHisSerLeuGluArGlnMetAsnGlyAsnGlnGlyLysG 985
||||| .....
1853 CCCACAGCGGAATTCGAGAGCTAGAGACATGCAAAATTAAGCCAGGAAG 1902
985 LyAspLysThr.....AspArGlyLysAspLys.....Thr 995
||||| .....
1903 TCAGAAAAAGCAAGAGAGAGAGCAAGAAAAAGCAAGGCGCAATTGAAA 1952
996 GlyLysGluLysLysLysAspAspAspLysGluLysAspLysMetLysAl 1012
||||| .....
1953 GTCAAGAGAGAAAAAGCCAAAGAGAGAGATGATCCAGAAAGAAAT 2002
1012 aLYsLYsGlyMetLeuLysGlyLeuGlyAspMetPheArGpHeGlyLysH 1029
||||| .....
2003 AAAGAACG.....AAGGCTTCGGCGCCCATCTGATGATTTGGAAAGA 2043
1029 tsArGlyAsp.....AspLysIleGluLysThrGlyLysIleLys 1042
||||| .....
2044 AGAAAGAGAGATTAAGCTGCAAGAGCTGAGCAAGAAAGTACTGTGAAA 2090

seq_name: /STD1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL08765
seq_documentation_block:
ID ABL08765 standard; cDNA; 5407 BP.
XX
AC ABL08765;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 20777.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NT.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
DR P-PDB: ABB64662.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
PS Claim 1; SEQ ID NO 20777; 21bp + Sequence listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent

```

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB16176-AB16175) and the encoded proteins  
 CC (AB5737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 5407 BP: 1382 A; 1436 C; 1482 G; 1107 T; 0 other;

alignment\_scores:                      Quality: 1171.50                      Length: 1656  
    Ratio: 1.396                                      Gaps: 68  
    Percent Similarity: 50.664                      Percent Identity: 26.932

alignment\_block:  
 US-09-757-781-2 x ABL08765                      ..

Align seg 1/1 to: ABL08765 from: 1 to: 5407

```

1 MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17
|||||
134 ATGAAGATCCGCTGCTGCTGCGACGATTCGCAATTCCTGCTGCTG6 183
17 yAspGlyHisMetLysValPheSerLeuIleGlnAlaValThrArgT 34
|||||
184 TTCGGCGCAACCTCCTCGCGCATCTCTCAAGAGCGCCAGCGCGCT 233
34 yTrArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
|||||
234 ACATTCAGAGCGCGCGCAAG...CCCGATTCTCGGGTACCGCGCGCAT 280
51 LeuGlnHisGlyAspGlyLysIleLeuAspLeuAspAspIleLeuCysAs 67
|||||
281 CTCGAG...ACACAGTCGGGCACTCTCGATCCGATGATGTCGCGCA 327
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 83
|||||
328 CGTGGCGGAGATCGGGAGAGCATATTTGGCGCATTTGATGACCCAGAG 377
84 .....ProHisHisGlyLysAspGlyThrSer...AlaSer 94
|||||
378 CCGATCCAGAGATTCGCAAGAGAGCGGAGATGAGCATCGGCGCATGCG 427
95 SerThrGlyThrGlnSerProGluIlePheGlySerGluLeuGlyThrAs 111
|||||
428 TCGGTTGGACCGGTTGCGCGCATCTTTCGCGATCC...ACCA 471
111 pAsnValSerAlaPheGlnProTyrGln...AlaThrSerGluIleGluV 127
|||||
472 CACGAGGCGCCCACTGCTCCGGGGATCTCCACGCCACACATCGAGG 521
127 alThrProSer.....ValLeuArgAlaAsnMetProLeuHis 139
|||||
522 TCACCCACACACATTCGGACCGATGGCTGATCGGATGGACTGATG 571
140 ValArgArgSerSerAspProAlaLeuIleGlyLeuSerThrSerValSe 156
|||||
572 GTGCGTCGACAGACTGATCCCAATCTCTGCGC...TCGCTGAAGAGAGA 618
156 rAspSerAspPheSerSerGluGluProSerArgLysAsnProThrArgT 173
|||||
619 GGGCAGGTAC.....AAAGCT 635
173 rPserThrThrAlaGlyPhe.....Leu 180
|||||
636 GGTGGGCGGCGCTCCCACTACGCTGGCGGAGATTCGCCGAGCGGCTG 685
181 LysGlnAsnThrAlaGlySerProLysThrCysAspArgLysLysAspG1 197
```

```

686 TTCTTGTACAAAGCCCGGTGGCCAGTTATGCCACAGGAGGAGAGACGA 735
197 uAsn.....TyrArgSerLeuProA 204
736 CGATCCCGCATCAGTACAGTAAAGAACAGCTGTCATCAACACACACCCC 785
204 rGAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAsnAlaArgSer 220
786 ATGCTGTAAATGAGAGCTCTCTCCGCTAACCATCAGCGTTGGCCGA 835
221 SerLeuSerAlaSerHisProMetValGly.....LysTr 232
836 TCCGGTCGCTGTCGATGACATTTCTTGGCGATGGCAATGCTACAGTG 885
232 pLeuGlnLysGlnLys.....GlnAspGluAspGlyThrGluGlu... 245
886 GATGAGGCGACGCGAAGAACTTCAGAAATCAGCCCGACCGCAGCAGCAT 935
245 ..... 245
936 ATCAGCAGGTTTCATCATCTGTCATAGTCAAAACGGTCTACTCC 985
246 .....AspAsnSerArgValGluProValGlyHisAl 256
986 AGCAAGTCTTCCCGAGGAGAGAGCAAGGAAAGAGCCCTTGGGACAGGC 1035
256 aAspThrGlyLeuGlnHisIleProAsnPheSerLeuAspAspMetValL 273
1036 ATATGATTCATCAGAGAG.....AAGATGTGTGAA 1067
273 yLeuValGluValProAsnAspGlyLysProLeuGlyIleHisValVal 289
1068 TGCTCTGATCATCAACGATACGCGCATCTCGCGGATCTGATCTG 1117
290 pProPheSerAlaArgGlyLysArgThrLeuGlyLeuValLysArgLe 306
1118 CCGGACAAAGAGCATGCGCGT.....GCAGCTGTGGTCCAAACATGT 1158
306 uGluLysGlyLysAlaGlnHisGlyAsnLeuPheArgLysGluAsnAspC 323
1159 GGAGCGCGGCGACCGAGCGAGAGAGAGAGACTG...CGTGGTGTGATC 1205
323 ySileValArgLysAsnAspGlyAspLeuArgAsnArgArgPheGluGln 339
1206 GCATTTGGAGATCAATGATCAAGTAAATGAGCTACACCGATCGGAG 1255
340 AlaGlnHisMetPheArgGlnAlaMetArgThrProIleIleTrpPheN1 356
1256 GTTCAGAGACACTCGCGCGGCTTGGAG..... 1285
356 sValValProAlaAlaAsnLysGluGlnTyrGluGlnLeuSerGlnSerG 373
1286 .....ACCTCAGAGTTGAGAGTGTGAGTGTGCGCGG 1319
373 LuLysAsnAsnTyrTyrSerSerArgPheSerProAspSerGlnTyrIle 389
1320 ATGCGAATCGC.....CGCAACAGCGT 1342
390 AspAsnArgSerValAsnSerAlaGlyLeuHisThrValGlnArgAlaPr 406
1343 GACTCTTAAGGTGGCCAGATGAGTGGAGTGGCTACCGTGTCA..... 1384
406 oArgLeuAsnHisProProGluGlnIleAspSerHisSerArgLeuProH 423
1384 ..... 1384
423 tSerAlaHisProSerGlyLysProProSerAlaProAlaSerAlaPro 439
1385 .....CCACCGCAACACATGCTGCTCGGAGGCGACACTCG 1423
440 GlnAsnValPheSerThrThrValSerSerGlyTyrAsnThrLysLysI1 456
|||||
```





```

860 .....
3247 GTTTCGGATATTAACATGCCAAGATCGATCCAGTATCGAGCAAGAG 860
860 luthrlyleuasn...ThrValAspaspGlnLysAlaGlySerProSer 875
3297 AGACGACAGCAATTTGGCCCTCGCGGAGCGGAGCGCAAGGATCAGCTG 875
876 ArgAspValGlyProSerLeuGlyLeuLysSerSerSerLeuGly 3346
3347 GCGCATCTGGGTCCATCGCTGGGAGCAAGAGAGTCTCTGCTGTAATC 892
892 lreuglnThrAlaValAlaGlnValThrLeuasnGlyAspIleProPhe 909
3397 GCTCCAGACTATGCTGAGAGCTGCAATGTC...GATAGCCGCTG 3443
909 lAsArgProArgProArgIleIleArgGlyArgGlyCysasnGlySerPhe 925
3444 GTCAACAGCGGTTCGCGCAACCGCGCTGCGCTGCGAGGAGACAGCTG 925
926 ArgAlaAlaAlaAspLysSerTyrAsp...LysPro... 936
3494 CGGCGCGGTGTGTCAGCAACCGATCGAGCAAGCCCGCAAGACCTG 3493
937 AlaValAspAspAspArgGlnGlyMetGlnThrLeuGlnGlnAspThrG 953
3544 GCTTTGGAGATGCGAT... 953
953 lGlnLysSerArgSerGlyArgGlnSerValSerThrAlaSerAspGln 969
3576 TTGCCTCGACGCCAATGGA... 969
970 ProSerHisSerLeuGlnLysGlnMetAsnGlnLysGlnLysGlnLys 3595
3596 ..... 3595
986 PylsthrAspArgLysLysAspLysThrGlyLysGlnLysLysAsp 3633
3634 CAAGTCGTGCGGGCCAG... 3632
1003 rGAspLysGlnLysAspLysMetLysAlaLysGlyMetLeuLysGly 3652
3653 ..... 3652
1020 LeuGlnLysPheThrPheArgPheGlyLysHisArgLysAspAsp 3673
3674 ATCGGTACATGTTCCGCTTGGCAAGATCGCAAGATGCGTGGTGGC 3723
1034 .....LysIleGlnLysThrGlyLysIleLysIleGlnGlnLys 3723
3724 AGTGATACATGTCGCTAAACATTTGCCCCCTACATCGGTGTTCCA 3773
1046 ePheThrSer... 3773
3774 CAGCCACATGCCCGCCAGCTGCACGACAGCAACACAGCAAGTTCACGAA 3823
1057 LysGlnGlnGlnLysGlnIleGlnAlaLysThrArgLysPheArgLys 1073
3824 CATACAGCAACMACAGATACCAACCGCTGATGGCTGCTGAGAG 3873
1073 g.GlnAlaArgLysArgAspTyrAlaGlnIleGlnAspPheHisArg 1089
3874 AAATGGCAAAACCGCGGCTATCAGC... 1089
1089 rPheGlyCysAspAspArgLysMetTyrGlyValSerSerTyrGlnG 3908
3909 CCGTCGCTGCTCC... 3908
1106 lYSerMetAlaLeuAsnAlaArgProGlnSerProArgGlnGlyHisMet 1122
3946 ..... 1122
1123 MetAspAlaLeuTyrAla...GlnValLysLysProArgSerSer 1137

```

---

```

3964 CAACCATCGCTATCCGATATTCACCAATTCAGAGAGCTCCACACAGAGC 4013
1137 sProSerProValAsp.SerAsnArgSerThrProSerAsnHisAspArg 1153
4014 ACCGACCAACCGAGATTACTGCCCC... 1153
1154 lIleGlnArgLeuArgGlnGlnLysPheGlnGlnAlaLysGlnAspGlnAsp 4048
4049 TACCATTG... 4048
1170 lGlnAspArgArgArgThrTyrSerPheGlnGlnLysProThrProAsnAla 1187
4084 G...AGCATGCACTGCAAGCAAGCTCCGATCCAGCCAGCAAGTCCCTGCTC 4130
1187 rGProAlaThrGlnSer...GlyArgHisSerValSer... 4130
4131 AGCCGCAAGCGCAATCAAAAGCGCTTCGCCCATGAGACAGTACTACAGAG 4180
1199 ...ValGlnValGlnMetGlnArgGlnGlnAlaArgGlnGlnLys 4180
4181 TAGGACAGGCTTCACAGCAAGCGGTGGGAGCATCAAGCATAGCCAGAG 4230
1214 rGlnGlnAlaGlnArgGlnLysTyrSerSerLeuProArgGlnSerArgLys 4231
4231 CAGCAGCGCC... 4231
1231 snAlaSerSerValSerGlnAspSerTyrGln... 4250
4251 CCTCCTGCCCATCAATGTCGCCGATTCGAAAGCGGCGCAGCCATTAAGGC 4300
1244 TyrSerProGlnGlyGlnLysPheGlnSerAlaLysGlnLysAsnProArgTyr 1260
4301 TACTGCGCACAGCCTGAAATAGAGTCCCG... 4333
1260 rSerTyrGlnGlnLysArgAsnGlyTyrLeuGlyGlyHisGlyPheAsn 1277
4334 ..... 1277
1277 lArgValMetLeuGlnLysGlnLysLeuLeuArgGlnLys 4351
4352 ..... 1290
1291 ..... 1291
4385 GCAATACCCGCCCACTGCTGCAGCAACATTCAGCAGCAAGTCCAGCA 4384
1300 sgInProProSer... 4434
4435 GCAGCAGCCCACTACAGACTTCACAGAAAGTCCAGCAATCCCAAT 4484
1310 rYAspSerTyrLysValAlaAsp.ProSerTyrAlaProProLysGln 1326
4485 ATGAGT... 1326
1326 yProPheArg 1329
4526 TGACTATGAG 4535

```

seq\_name: /STD1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.ABL08764

seq\_documentation\_block:

ID ABL08764 standard; cDNA: 3636 bp.

AC ABL08764;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 20774.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.





```

783 spAspAlaGlyPheValThrAlaSpAlaGlyThrTrpAlaLysAlaAla 799
    |||.....|
31874 ATGGCAGCAGCTCAACACAGCAGCTGCTAATCATCGCAGAACTGGCC 31923
800 Ile..... 800
31924 GTGGCAGATCAGGGCCACCATAAATACGTCCTGGTACTCCGTGCAC 31973
801 SerAspSerAlaSpCysSerLeuSerProAspVal..... 812
31974 ATCCAGCAACTTGTGATGCCACTACTCTCTGCAACTGACTGTGAGCA 32023
813 AspProValLeuAlaPheGlnArgGluGlyPheGlyArgGlnSerMet 828
    |||.....|
32024 ACTCGGGCGTGAGACATTTCTCGCGCATGCTTTGGAGCAGCAGCATC 32073
829 SerGluLys..... 831
32074 TCTGAGAAAGCAGCATGCGCGCTGGATGCCGGAAACTGCACTATCA 32123
832 ArgThrLysGlnPheSerAspAlaSerGlnLeuAsp...PheValLysT 847
    |||.....|
32124 GCGGATTAAGAACTTACGCGAGGAGCGCAACGCGAGCTCGCATTCAGC 32173
847 hrArgLysSer.....LysSerMetAspLeuGly 856
    |||.....|
32174 TGACCAATCGCGTGTATGTGTATGTCATGATCCCTTACGCGCTCGC 32223
857 IleAlaAsp..... 859
32224 ATAGCCAGTGTCTAATGCCAGTTTGGGATATTAACATGCCAAGACTGC 32273
860 .....GluThrLysLeuAsn...ThrValAspArg 869
    |||.....|
32274 ATCCAGATATGAGCAAGAGAGAGCAGCAATGTGGCGGTGGAGAG 32323
869 LysAlaGlySerProSerArgAspValGlyProSerLeuGlyLeuLys 885
    |||.....|
32324 CGAGGCGAAGGATCAGCTGCGGATCTGGTGCATCGCTGCGGATGAG 32373
886 LysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluValThrLe 902
    |||.....|
32374 AATCTCTCGTGTGAGTGTGCTGCACACTATGTCGAGGCTGCAGAT 32423
902 uAsnGlyAspIleProPheHisArgProArgIleIleArgGlyA 919
    |||.....|
32424 GTGG...GATGAGCCCGGTGTGCATCAGCGGTGCGGCAACGCGTGGC 32470
919 rGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerTyrAspLys 935
    |||.....|
32471 GTGCGAGGAGGAGCAGCTGCGGCGCGGTGTCACGCAACGAGATGCG 32520
936 ProAla.....ValAspAspAspGluGlyMetGlu..... 946
    |||.....|
32521 AGCAATGATGTTTGGATATAAACAAGATTTCTGTGAATGTGAATAT 32570
947 .....ThrLeuGlnG 950
    |||.....|
32571 TTGTATGCGCTATATCTTTTATAGACCCGCAAGACCTGCTTTTGAGG 32620
950 LysAspThrGluGlu.....SerSerArgSerGlyArgGlnSer 962
    |||.....|
32621 ATGCGCATCAGAGGTGTTTGGCTTCGCGCAGCGCAATGGA..... 32661
963 ValSerThrLysSerAspGlnProSerHisSerLeuGlnLysGlnMetAs 979
    |||.....|
32662 .....CCGTTCCAGAGTTCCCTCGAA 32681
979 nGlyAsnGlnGluGlyAspLysThrAspArgLysLysAspLysThrG 996
    |||.....|
32682 CGATGGCAAAACAC...GGCTGCAAGTGTGCGGCGCAAG..... 32718

```

```

996 LysGluLysLysLysAspArgAspLysGluLysAspLysMetLysAla 1012
32718 ..... 32718
1013 LysLysGlyMetLeuLysGlyLeuGlyAspMetPheArgPheGlyLysH 1029
    |||.....|
32719 AAGCCAAGCATACTGCGCGGTATGCTGACATGTTCCGCTTGGCAAGAA 32768
1029 sarGlyAspAsp.....LysIleGluLysThrG 1039
    |||.....|
32769 TCGCAAGATATGCGCTGTGCGCAGTGAATTAATGCGGTAAACATTTCCG 32818
1039 LysLysIleLysIleGlnLysSerPheThrSer..... 1049
    |||.....|
32819 CCCCTACATGCGGTGTTCCACAGCCACATGCGCGCAGCTCAGCAGAG 32868
1050 GlnGluLysArgIleArgMetLysGlnGlnLysGlnLysGlnLysAla 1066
    |||.....|
32869 CACAGCAGCAGTTGACAGCAATCAGCAGCAGCAGCAGATGATCCAGCCG 32918
1066 sThrArgGluPheArgGluArg.GlnAlaArgGluArgAspTyrAlaGlu 1082
    |||.....|
32919 TGCATTTGCTGCTCTGTGAGAGAAATGCGCAACCGCGCGTATCAGC... 32965
1083 IleGlnAspPheHisArg.ThrPheGlyCysAspAspGluLeuMetLys 1099
    |||.....|
32966 .....CACCCACCGCTGCTGCTC.....AAATGGA 32994
1099 LysGlyValSerSerTyrGlnGlySerMetAlaLeuAsnAlaArgProGln 1115
    |||.....|
32995 GTGGTATGCAATGGTAT..... 33011
1116 SerProArgGlnGlyHisMetMetAspAlaLeuTyrAla.....GlnVa 1130
    |||.....|
33012 ...CCATCAGACAGCAGATCTTCAACATTCGCTATACAGATTAATGCCACT 33058
1130 LysLysProArgAsnSerLysProSerProValAspSerAsnArgSert 1147
    |||.....|
33059 ACGAGGACCTTCACACAGCAGCAGCAGCAACACAGATTAGGTGAGAGCGG 33108
1147 hrPro 1148
    |||.....|
33109 TCCCG 33113
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH07109
seq_documentation_block:
ID AAH07109 standard; cDNA; 556 BP.
XX
XX AAH07109;
XX
XX AC
XX 26-JUN-2001 (first entry)
XX
XX DE Human cDNA clone (5'-primer) SEQ ID NO:3944.
XX
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX OS Homo sapiens.
XX
XX PN EP1074617-A2.
XX
XX PD 07-FEB-2001.
XX
XX PF 28-JUL-2000; 2000EP-0116126.
XX
XX PR 29-JUL-1999; 99JP-0248036.
XX
XX PR 27-AUG-1999; 99JP-0300253.
XX
XX PR 11-JAN-2000; 2000JP-0118776.
XX
XX PR 02-MAY-2000; 2000JP-0183767.
XX
XX PR 09-JUN-2000; 2000JP-0241899.
XX
XX PA (HELI-) HELIX RES INST.
XX

```

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-18749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -

PS Claim 1, SEQ ID 3944; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification, where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and  
CC AAH95893 represent human amino acid sequences; AAH92446 to  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 556 BP; 154 A; 129 C; 144 G; 126 T; 3 other;

# alignment\_scores:

Quality: 936.00 Length: 184  
Ratio: 5.200 Gaps: 0  
Percent Similarity: 97.826 Percent Identity: 97.283

# alignment\_block:

US-09-757-781-2 x AAH07109

Align seg 1/1 to: AAH07109 from: 1 to: 556

126 GIUValThrProSerValLeuAlaAsnMetProLeuHisValArgAr 142  
|||  
3 GAGGTACACCTTCAGTCTTCGAGCAAAATATGCTCTTCATGTTGACG 52  
142 gserSerSprAlaLeuIleGlyLeuSerThrSerValSerAspSera 159  
53 CAGTAGTACCCAGCTCTAATTCGCTTCACCTGTCACGTGATAGTA 102  
159 snpSerSerGluGluProSerArgValAsnProThrArgTrpSerThr 175  
103 ATTTTCTCTGAAGACCTTCAGAGGAAAAATCCACACGCTGGTCAACA 152  
176 ThrAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysTrpCysAs 192  
153 ACAGCTGGCTCTCCACAGCAGAACCTCTGCGAGTCCCTGAGTCCAGT 202  
192 PArgLysLysAspGluAsnTyrArgSerLeuProAlaArgAspThrSerAsnT 209  
203 CAGGAGAAAGATGAAACCTACAGAACCTCCCGGAGATAGTAAGTAACT 252  
209 rFserAsnGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSer 225  
253 GGCTTAACCAATTTCAGAGACAAATGCTGCTGCTCTGAGTCCAGT 302  
226 HisProMetValGlyLysTrpLeuGluLysGlnGluAspGluAspGlu 242

|||||  
303 CACCAATGTGGCAAGTGGCAGGAACAGACAGATGAGATG 352  
242 YThrGluGluAspAsnSerArgValGluProValGlyHisAlaAspThrG 259  
353 GACGAGAAAGAGATTAACAGTCTGTTGAACCTTGGACATGCTGCACACG 402  
259 LyeuGluHisIleProAsnPheserLeuAspAspMetValLysLeuVal 275  
403 GTTGGAGCATATACCAACTTTCTCTGATGATATGTAAGCTCGTA 452  
276 GIUValProAsnAspGlyLysProLeuGlyLeHisValAlaProPheSe 292  
453 GAAGTCCCAACGATGAGAGGCTCTGGAAATCCATGATGTCCTTTGAG 502  
292 rAlaArgGlyLysArgPheLeuGlyLeuValLysArgLeuGluLysG 309  
503 TGTTCGAGNCGGCANNAACCTGGGTTATAGTAAACGATTGGGAAAG 552  
309 LY 309  
553 GT 554  
seq\_name: /SID51/gcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH71480  
seq\_documentation\_block:  
ID AAH71480 standard; cDNA: 597 BP.  
XX AAH71480;  
XX  
XX 19-SEP-2001 (first entry)  
DE Human cervical cancer marker nucleic acid 2754.  
XX  
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX Homo sapiens.  
XX  
XX WO200142467-A2.  
XX  
XX 14-JUN-2001.  
XX  
XX 08-DEC-2000; 2000MO-US33312.  
XX  
XX 08-DEC-1999; 99US-0169681.  
XX 21-DEC-1999; 99US-0171350.  
XX 14-MAR-2000; 2000US-0189315.  
XX 12-MAY-2000; 2000US-0203791.  
XX 09-JUN-2000; 2000US-0210600.  
XX 21-JUL-2000; 2000US-0220114.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Schlegel R, Deeds J, Berger A, Zhao X;  
PI WPI; 2001-375006/39.  
XX  
XX New isolated nucleic acid for diagnosing and treating cervical cancer  
PT and for assessing and detecting compounds for treating the cancer -  
PT  
XX Claim 1; Page 566; 1051pp; English.  
XX  
XX The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC progression of cervical cancer or has a pre-malignant condition; to monitor the  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
XX useful for gene therapy.  
XX  
XX Sequence 597 BP; 180 A; 123 C; 156 G; 138 T; 0 other;

## alignment\_scores:

Quality: 902.50 Length: 209  
Ratio: 4.725 Gaps: 1  
Percent Similarity: 91.388 Percent Identity: 88.995

## alignment\_block:

US-09-757-781-2 x AAH71480 ..

Align seg 1/1 to: AAH71480 from: 1 to: 597

```
468 ThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyI 484
|||||
1 ACAGAAAGGTTGGGATTGACATCACCACGAGATGTAAACAATAGGTGG 50
|||||
484 YSerAlaProIleThrValLysAsnIleLeuProArgGlyAlaAlaIleG 501
|||||
51 CTACAGCTCCAAATCTATGTGTAACAAATCTCCCGGGGGGGCCCAATC 100
|||||
501 LAspGlyArgLeuLysAlaGlyAspArgLeuIleGlyValaAsnGlyVal 517
|||||
101 AGCATGGCCGACTTAAGCGCAGACGAGACTATAGAGTAAATGGAGTA 150
|||||
518 AspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerTh 534
|||||
151 GATTAGTGGGCAATCCCAAGAGAGATTGTTGCTGTTGAGAAACAC 200
|||||
534 rLysMetGluGlyThrValSerLeuValPheArgGlnGluAspAlaP 551
|||||
201 CAAGATGGAAAGAACTGTGAGCTTCTGCTTCCGCCAGAAAGACGCT 250
|||||
551 heHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLys 567
|||||
251 TCCACCCCAAGGAACGTG..... 267
|||||
568 GluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrAr 584
|||||
268 .....AAACGAGAAAGATGAGATTGTTCTTACACCTGATGGCACACAG 311
|||||
584 gGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGlyL 601
|||||
312 GGAATTCTGACATTTGAACTCCCACTTAATGATCAAGATCTGCAGGCC 361
|||||
601 euGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLeu 617
|||||
362 TTGGTGTCAAGTCAAAAGGTAAACCGGTCAAAAGAGAACCAACGAGTTTG 411
|||||
618 GlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
|||||
412 GGAATCTTTGTCAAGTCCATTATTAAAGGAGGCGACCATCTAAAGATGG 461
|||||
634 YArgLeuArgValAsnAspGlnLeuIleAlaValaAsnGlyGluSerLeu 651
|||||
462 AAGGCTTTGGTGAATGATCACTGATAGCAATTAATGAGAAATCCCTGT 511
|||||
651 euGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSer 667
|||||
512 TGGGCAAGAACCAACCAAGATGCCCTGGAACCAACCTAAGAGCTATGTCT 561
|||||
668 ThrGluGlyAsnLysArgGlyMetIle 676
|||||
562 ACTTGAGGCCCATTAACGAAAGATGATC 588
```







```

/ gene="PAR03"
/ note="Region: PDZ domain"
misc_feature / gene="PAR03"
/ note="Region: PDZ domain"
BASE COUNT 1220 a 923 c 1079 g 849 t
ORIGIN

alignment_scores:      Length: 1356
Quality: 7035.00      Gaps: 0
Ratio: 5.188
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:
US-09-757-781-2 x AF467002 ..

Align seg 1/1 to: AF467002 from: 1 to: 4071

1 MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17
|||||
1 ATGAAGTACCGTGTCTCGAGCGAGACCGGGTGGTGGTGGTGGTGG 50
|||||
17 yAspGlyHisMetLysValPheSerLeu11egInGlnAlaValThrArg 34
|||||
51 GAGCGCCACATGAAAGTTTCAGCTTCATCCACAGCGCGGTGACCCG 100
|||||
34 yTrArgLysAlaIleAlaLysAspProAsnYrTrpIleGlnValHisArg 50
|||||
101 ACCGAGAGCGCATGCCAAGATGCCAAGCTAGTGAATACAGTGCATCGC 150
|||||
51 LeuGlnHisGlyAspGlyLysIleLeuAsnLysAspIleLeuCysAs 67
|||||
151 TTGGAACATGAGATGAGAGAACTAGACCTGTGATGACATTTCTTGTGA 200
|||||
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnAspP 84
|||||
201 TGTAGACAGCATTAAGACAGACTGTAGCAGTGTGATGAGCAGATC 250
|||||
84 rOHSHISGlyLysPglyYrThrSerAlaSerSerThrGlyThrGlnSer 100
|||||
251 CACATCAGGAGTGTATGGCACCGATGCCAGTCCAGCGGTACCCAGAC 300
|||||
101 ProGluIlePheGlySerGluLeuGlyYrThrAsnAsnValSerAlaPheG1 117
|||||
301 CCAGAGATATTGGTAGTGAAGCTTGGCACCAACAAATGTCTACGCTTTCA 350
|||||
117 nProTyrglnAlaThrSerGluIleGlnValThrProSerValLeuArgA 134
|||||
351 GCCTTACCAAGCACAGTGAATTTGAGGTACACACTTCAGTCCCTTCGAG 400
|||||
134 lAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
|||||
401 CAAATTTGCTCTTCATGTTGCAGCCAGTAGTACCCAGCTCTAATTGGCC 450
|||||
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluIleProSerAr 167
|||||
451 CTCTCCACTTCTGTCAAGTATATTTTCTCTGAAGACCTTCAG 500
|||||
167 gLysAspProThrArgTrpSerThrAlaGlyPheLeuLysGlnAsnTr 184
|||||
501 GAAAAATCCACACAGCTGTCAACACACAGCTGCTTCCCTCAAGCAACA 550
|||||
184 nAlaGlySerProLysThrCysAspArgLysLysAspGlnAsnYrArg 200
|||||
551 CTGCTGGAGATCTCTAAACCTGCGACAGAGAAAGTGAACCTACACA 600
|||||
201 SerLeuProArgAspPThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217
|||||
601 AGCTTCGCCGCGGATACCTAGTACCTGGTCTTACCAATTTACAGAGACA 650
|||||
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeu 234
|||||

```

```

651 TGTGCTGCTCTCTAGTGCAGTCCACCCATGTTGGCGCAAGTGGCTG 700
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
|||||
701 AGAACAAGAACAGATGAGATGGAGACAGAGAGATTAACAGTGGTGT 750
|||||
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
|||||
751 GAACCTGTGGACATGCTGACAGGGGTTGGAGCATATACCACTTTC 800
|||||
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284
|||||
801 TGTGATGATATGTGTAAGCTCTAGAAAGTCCCAACGATGAGAGGCCCTC 850
|||||
284 euGlyLysHisValValProPheSerAlaArgGlyLysArgThrLeuGly 300
|||||
851 TGGAAATCCATGTAGTGCTTTCAGTCTCGAGGCGGCAAGACCTCGGG 900
|||||
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisGlnAsnLe 317
|||||
901 TTTATTGTAAGCATGTGAGAAAGGTGTAAAGCTGAAACATGAATAATCT 950
|||||
317 uPheArgLysAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
|||||
951 TTTTCGTGGAAGTATGATGCTTCAGGATTAATGATGCCACCTTGAA 1000
|||||
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
|||||
1001 ATGAAAGATTTGAAACAGCACAAACATATGTTCCGCAAGCCATGGTACA 1050
|||||
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnYrG1 367
|||||
1051 CCGATCATTTTGTTCCATGTGGTCTCGACCAAAATAAGACCATATAGA 1100
|||||
367 uGlnLeuSerGlnSerGluLysAsnAsnYrYrSerSerArgPheSerP 384
|||||
1101 ACAATATCCCAAGTAGAGAAACAAATTAATTCATACCGCTTTTAGCC 1150
|||||
384 rAspSerGlnYrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
|||||
1151 CTGACAGCCAGTATATTGACAAAGAGGTGAACTGACAGGGCTTCA 1200
|||||
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
|||||
1201 ACGGTCCAGAGAGACCCGACTGAAACACCCGCTGAGCAGTAAAGCTC 1250
|||||
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
|||||
1251 TCACCTAAGACTACCTCATAGCGCACACCCCTCGGGAACACCATCCG 1300
|||||
434 lAspAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
|||||
1301 CTCGAGCTCGGCACTCAGAAATGTTTATGACACTGTAAAGCTGTGT 1350
|||||
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysG1 467
|||||
1351 TATAACACCAAAAATATGAGCAAGAGCTTAATATTCACACTTAAGAAAG 1400
|||||
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484
|||||
1401 TACAGAAAGTTTGGGATTCACACATCACTTCAGAGATGTAACATAGGTG 1450
|||||
484 lYSerAlaProIleYrValLysAsnIleLeuProArgGlyAlaAlaIle 500
|||||
1451 GCTCAGCTCCAAATCTATGTGAAACATTTCTCCCGGGGGGCGGCATY 1500
|||||
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGlnValAsnGlyVa 517
|||||
1501 CAGGATGCCGCACTTAAGCAGAGACAGACTTATAGAGTAAATGAGAT 1550
|||||
517 lAspLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSerT 534
|||||
1551 AGATTAGTGGCAAAATCCAAAGAGAAATGTTTCCGCTGTGAGAAACA 1600
|||||

```

534 hrLysMetGluGlyThrValSerLeuLeuValPheArgInGluAspAla 550  
|||||  
1601 CCAAGATGAGGAAGACTGTGAGCCTTCTGTCTTTCGCCAGGAAGACGCC 1650  
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnLeuProly 567  
|||||  
1651 TTCCACCCCAAGGAACTGAATCGAGCCAGCCAGCATCGAGATTCCAAA 1700  
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrA 584  
|||||  
1701 ACAAAGCAAGCAAGAGATGAGGATATGTCTTACACCTGATGGCAGCA 1750  
584 rGluLpheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600  
|||||  
1751 GGGAAATTCTGACATTTGAAGTCCCACTTAATGATTCAGGATCTGCAGG 1800  
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617  
|||||  
1801 CTTCGTGTAGTGTCAAGTAACCGGTCAAAAGCAACACCAGCATTTT 1850  
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaIleAspLysAspG 634  
|||||  
1851 GGGAACTTTTGTCAAGTCCCATTTATATGAGAGAGCAGCATCTAAAGATG 1900  
634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650  
|||||  
1901 GAAGGCTTGGGGTGAATGATCAACTAGTAGCAGTAATGAGAGATCCCG 1950  
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667  
|||||  
1951 TTGGGGCAACAACAACCMAGATGCGTAAGACCTTAAGAAAGCTATATGTC 2000  
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684  
|||||  
2001 TACTGAAGCATAAACAAGAGATGATCCAGCTTATTTGTCAGAGAGAA 2050  
684 LeuSerLysCysAsnGluLeuLysSerProGlySerProGlyProGlu 700  
|||||  
2051 TAAGCAAGTGCATAGTAGTGAAGTCACTGGGAGCCCTCGACCTGAG 2100  
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717  
|||||  
2101 CTGCCCATTTGAAACACCGTTTGATGATAGAGAAAGAAATTTCCCATTC 2150  
717 rLeuThrSerGlyIleGluGlyLeuAspGluSerProSerArgAspAlaA 734  
|||||  
2151 CCTCTACAGTGGGATTTGAGGGGCTTATGATATCCCCAGCAGAAATGCTG 2200  
734 LeuSerArgIleMetGlyGluSerGlyLysThrGlnLeuSerProThr 750  
|||||  
2201 CCTCATGATGGATTAATGGGTGAGTCAAGTAAATACAGCGTCTCCCTACA 2250  
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767  
|||||  
2251 GTCAAAATATGCCCCAAGATGACACTGTCAATTATGAAAGATGACAGTGGCC 2300  
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerSerHisAspA 784  
|||||  
2301 AGTGCTTCCTCCACATCTCTGACCAAGTCTCTCCAGCTCCACAGATG 2350  
784 sPValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800  
|||||  
2351 ATGTGGGTTTGTGACGCGAGATGCTGTACTTGGGCCAAAGCTGCATATC 2400  
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817  
|||||  
2401 AGGTATTCAGCCGACATGCTTTGAGTCCAGATGTGATCCAGTCTTCGTC 2450  
817 aPheGlnArgGluGlyPheGlyLysArgLysMetSerGluLysArgThrL 834  
|||||  
2451 TTTTCAACGAAAGAAATTGGACGTCAAGATATGTCAGAAAAACGACAA 2500

834 ySgInPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850  
|||||  
2501 AGCAATTTTCAGATGCCAGATCAATTTGATTTGTTAAACACAGAAATCA 2550  
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867  
|||||  
2551 AAAAGATGATTTTATGATTAAGCTGACGAGACATTAACCTCAATACAGGGA 2600  
867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884  
|||||  
2601 TGACCAAGAAAGCAGGTTCTCCAGCAGAGATGTGGGTCTTCTCCCTGGGTC 2650  
884 eLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900  
|||||  
2651 TGAAGAAGTCAAGCTCGTTGAGAGAGTCTGCAGACCGCAGTTGCCAGAGTG 2700  
901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleArg 917  
|||||  
2701 ACTTTGAATGGGATATTCCTTCCATCGTCCACGGCCGGGATTAATCAG 2750  
917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyra 934  
|||||  
2751 AGCCAGGGGATGCCATGAGAGCTTCAGAGCTGCCATGCAGAAATCTTATG 2800  
934 sPlysProAlaValAspAspAspAspGluGlyMetGluThrLeuGluGlu 950  
|||||  
2801 ATAAACCCCGGTAGATGATGATGATGATGATGATGATGATGATGATGATG 2850  
951 AspThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlaSe 967  
|||||  
2851 GACACAGAAAGAAAGTTCAATCATGAGGAGAGAGCTGTATTCACACAGCAG 2900  
967 rAspGlnProSerHisSerLeuGluArgGlnMetAsnGlyAsnGlnGlu 984  
|||||  
2901 TGATCAAGCTTCCCACTCTCTGGAGAGACAAATGAATGAAGAACCAAGAGA 2950  
984 ySgIysPlysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000  
|||||  
2951 AAGGTATTAAGACTGATAGAAATAAAGATTAACCTGGAATAAAGAAAGAG 3000  
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017  
|||||  
3001 AAAGATTAAGATTAAGAGAAAGATTAATAAAGCCAAAGAAAGGAAAGCT 3050  
1017 uLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspL 1034  
|||||  
3051 GAAGGGCTTGGAGACATGTTCAGGTTTGGCAACATCGAAAGATGATACA 3100  
1034 ySgIleGluLysThrGlyLysIleLysIleGlnGlnLeuSerPheThrSerGlu 1050  
|||||  
3101 AGATTGGAATAACGGGTAAATAAATAATACAGGAATCTTTACATCGAA 3150  
1051 GluGluArgIleArgMetLysGlnGlnGlnGluArgIleGlnAlaLysTh 1067  
|||||  
3151 GAGGAGAGGATAGCAAAAGAAAGAGAGAGAGAGAGAGATTCACAGCAAAAC 3200  
1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyraIleGluIleG 1084  
|||||  
3201 TCAGAAATTTAAGGAAAGCAGACAGCTCGAGAGCTGAGCATATGCTGAATTC 3250  
1084 InAspPheHisArgThrPheGlyCysAspAspGluLeuMetTyrglyGly 1100  
|||||  
3251 AAGATTTTCATCGGACATTTGGCTGTGATGATGATGATTAATGTATGGGGA 3300  
1101 ValSerSerTyrgluGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117  
|||||  
3301 GTTTCCTTTATGAAAGGTTTCATGCGCTCTCAACGCTAGACCTCAGAGCCC 3350  
1117 oArgGluGlyHisMetMetAspAlaLeuTyraIleValLysLysPro 1134  
|||||  
3351 ACGAGAAAGGCAATATGATGATGATGATGATGATGATGATGATGATGATG 3400  
1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150

seq_name	gb_pr	AF467003
seq_documentation block:		
LOCUS	AF467003	4062 bp mRNA linear PRI 07-FEB-2002
DEFINITION	Homo sapiens partitioning-defective 3 protein splice variant b (PAR3D), mRNA, complete cds, alternatively spliced.	
ACCESSION	AF467003	
VERSION	AF467003.1	GI:18568345
KEYWORDS		
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 4062)	
TITLE	Gao L., Macara, I.G. and Joberdy G.	
JOURNAL	Multiple splice variants of PAR3 and of a novel related gene, PAR3L, produce functionally different proteins	
Unpublished		

2 (bases 1 to 4062)  
 AUTHORS Gaol, L., Macara, I. G. and Joberty, G.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-JAN-2002) Center for Cell Signaling, University of  
 Virginia, PO Box 800577, Charlottesville, VA 22908, USA

FEATURES  
 source  
 1..4062  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..4062  
 /gene="PARD3"  
 /note="PAR3"  
 1..4062  
 /gene="PARD3"  
 /note="PARB; multi-PDZ domain containing protein;  
 alternatively spliced"  
 /codon\_start=1  
 /product="partitioning-defective 3 protein splice variant  
 b"

CDS  
 /protein\_id="AA176043.1"  
 /db\_xref="gi:18568346"  
 /translation="MKYVVCGRTRVYPCGDMKMYFSLIQAVTRYRAIAKDPY  
 WIQVRLHEHGGGLIDLDLIDVADKDRLVAVFDQDPHGGDGTASSTGQSP  
 IFGSELGTNNYSAFOYQATSEIVTSVLRANPDLHRRSSPALIGLSTVSDSNF  
 SSESSEKRNPTRMSTAGTGLKONTAGSPKTCRKKDNRSLPDTGMSNPORONA  
 RSTSSASHPMYKMLEKOEDEDGTEEDNSRVEVGHADGLEHIIPEMSLDWKIVE  
 VPNGQPLGIIHVPEPSARGGRTGLIKRLKLGKAKHEMLFRNCTIVRIINODLNR  
 RRPQAOAHMRQARTRPIIMPHVYPAANKRQYEDLSQSEKNNYTSRFPSPQITDKR  
 SVNSAGLHTVORAPRLNHPPEQIDSHRSLPHSAHPSCKPPSAPASAPQNVFTVSSG  
 YNRKIKGRNLNIDLRKTEGLGFSITSRDVLTGSAPITYKNILPRKALIDGRLKAG  
 DRLELVNGVLDVGRKSQEEVYSLKSTMEGTSLVLRQDAFPIRLNADPEPQMQIP  
 KETAEDEDIVLAPDGTREFTEVEPLNGDSAGLGVSVGNRSKENHADLGIVKSI  
 INGSAASKDRLARVNDQILIANVNGSLIGKINODAMELTRSMSTEGKRMQITLVAR  
 RISCNELKSGSPGPELPLETALDDREIRISHLSVSGIEGLDESPRNAALSRING  
 KYOLSPVNMPODDTVIIEDDRLLPVLIPPHISDQSSSHDQGVPAADACTWAKAIS  
 DSNACSLSPYVDVPLAFOREBERQSSSEKTRKQFSDASQIDTVYTKKSSMDGLAD  
 ETKNATVDDQKAGSPKSDVGSJGLKSSLSLEQTVAAEVTNLNGDIPFHRPRILIR  
 GRCNTEFRAIDKSYDKPAVDDDEGEMLEDETESSRSRGESVSTADSDSHLE  
 RMNGNOMEKDKTDRKDKTKGKKKDKDRKEMKAKMGLGIDMPFRFGHRKD  
 KIEYTKGIKIOESTSEERIIMKQEQERIOAKTREFRQAREDAVETIDFPRFG  
 CDDILMAGVSYGSMALNAPQSPREGHMALVYQVKKPNRSKPSVDNSRTPS  
 NHDHIDRLROEFOYAKODEVEDRRRTYSPQMPNARVATQGRHSVSYEQVQR  
 QERESQGNORQYSSIPROSKRAASSVSDSMQNTSPEDGTSKAEPRYSYIQSS  
 RNGLGEGFNARVMELEQLROEQKRRKQMKQKPPSEGPSNYDSYKRVQDPSTAP  
 PKGFGRODVPSPSPQVARNRLQIPKGRFPYS"  
 814..1074  
 /gene="PARD3"  
 /note="Region: PDZ domain"  
 1378..1638  
 /gene="PARD3"  
 /note="Region: PDZ domain"  
 1765..2028  
 /gene="PARD3"  
 /note="Region: PDZ domain"  
 922 c 1075 g 847 t

BASE COUNT 1218 a  
 ORIGIN

alignment\_scores:  
 Quality: 7008.50 Length: 1356  
 Ratio: 5.180 Gaps: 1  
 Percent Similarity: 99.779 Percent Identity: 99.779

alignment block:  
 US-09-757-781-2 x AF467003 ..

Align seg 1/1 to: AF467003 from: 1 to: 4062

1 MetLysValThrValCysPheGlyIArgThrArgValAlaValProCysG1 17  
 |||||  
 1 ATGAAGAAGTACCGTGTGCTTCGACGAGACCGGAGTGCGTGCCTCGCG 50  
 |||||  
 17 yAspGlyHMetLysValAlaPheSerLeuIleGlnIleAlaValThrArgT 34

```

|||||
51 GGACGGCCACATGAAAGTTTTCAGCCTCATCCAGACGGCGTACCCGCT 100
34 yfArqYsAlaIleAlaYsAPProAsnTyrTrpIleGlnValHisArg 50
|||||
101 ACCGGAAGCCATCCGCCAAGATCCAAACTACTGATACAGTGCCATCCG 150
51 LeuGlnHisGlyAspGlyLysIleLeuAspLeuAspAspIleLeuGlyAs 67
|||||
151 TTGGACATGTGAGATGAGGAATACTAGACCTTGATGACATCTTGTTGA 200
67 pValAlaAspAspLysAspArgLeuValAlaValAlaPheAspGluGlnAsp 84
|||||
201 TGTAGAGAGAGATAAAGACAGACTGTGACGTGTTGATGAGACAGATC 250
84 roHisHisGlyLysAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
|||||
251 CACATCACGAGAGTGATGGCACCATGTCCTCCACGGGTACCCAGAGC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG 117
|||||
301 CCAGAGATATTGTGTAGTGTGGCACCAACAATGTCACAGCTTTTCA 350
117 hProTyrGlnAlaThrSerGluIleGlnValThrProSerValLeuArg 134
|||||
351 GCCTTACCAAGCAACAAGTGAATTGAGGTCAACCTTCAGTCCCTCGAG 400
134 laAsnMetProLeuHisValArgArgSerSerProAlaLeuIleGly 150
|||||
401 CAAATATGCCCTTCATGTTGACGCAAGTAGTACCACCGCTTAATTGGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
|||||
451 CTCTCCACTCTGTCAGTGATTAATTTTCTCTGGAAGAGCCTTCAAG 500
167 glysAsnProThrArgTyrSerThrThrAlaGlyPheLeuLysGlnAsn 184
501 GAAAAATCCACACGCTGTCAACAACAGCTGCTTCCTCAACACAGACA 550
184 hValAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
|||||
551 CTGCTGGAGACTCTAAACCTCGCACAGAGAAGAAAGTGAACACTCAGAG 600
201 SerLeuProArgAspThrSerAsnTyrSerAsnGlnPheGlnArgAsp 217
601 AGCTCCCGGGGATAGTACTAGTAACTGGTCAACCAATTCAGAGAGACA 650
217 naIlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTyrLeu 234
|||||
651 TGCTCCCTGCTCTGAGTGCAGTCAACCCAATGGTGGCAAGTGGCTGG 700
234 lIuysGlnGluGlnAspGluAspGlyThrGluGlnAspAsnSerArgVal 250
|||||
701 AGAAACAGAGACAGATGAGGATGGACAGACAAGATTAACACTCGTGT 750
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
751 GAACGTGTGACATGCTGACACGGGTTTGGACATATACCACAACCTTTC 800
267 rLeuAspAspMetValLysLeuValGlnValAlaProAsnAspGlyLys 284
|||||
801 TCTGTGATGATGTAAAGCTCTAGAAAGTCCCAACAGATGAGAGGCTTC 850
284 euGlyIleHisValAlaProPheSerAlaArgGlyLysArgThrLeuGly 300
|||||
851 TGGGATCCCATGTAGTCCCTTTCAGTCCCTCGAGCGGCGACAGACCTTGG 900
301 LeuLeuValLysArgLeuGlyLysGlyLysAlaGluHisGluAsnLe 317
901 TTATTGTAAACGATTTGAGAAAGGTGTAAAGCTGAACATGAAAAATCT 950
317 upHeArgGluAsnAspCysTlleValArgIleAsnAspGlyAspLeuArg 334
|||||
951 TTTTCTGAGAAATGATTCATGTCTCAGATTAATGATGGCGACCTTCGAA 1000
334 snArqArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
|||||
1001 ATAGAAATTTGAACAAGCAACAACATATGTTTGGCCAAAGCCATGCGTACA 1050
351 ProIleIleThrPheHisValAlaProAlaAlaAsnLysGluGlnTyrG 367
|||||
1051 CCCATCTTTGGTTCATGTGTGTTCTTCGACGAAATTAAGACAGATGTGA 1100
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSer 384
1101 ACAACTATCCCAAGGTGAGAAACAATTAATTAATCAAGCCGTTTAAGCC 1150
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
|||||
1151 CTGACAGCCAGTATATTGACACAGAGTGTGAACAGTGCAGGGCTTCAC 1200
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
1201 ACGGTGACAGAGACACCCGACGTGAACACCACCGCTGACAGATAGACTC 1250
417 rHisSerArgLeuProHisSerSerAlaHisProSerGlyLysProProSer 434
1251 TCACTCAAGACTACCTCATAGCGCACACCCCTCGGGAACACCATCCG 1300
434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1301 CTCACGCTTCGCACTTCAGAAATGATTATTTAGTACAGCTGTAAGCAGTGT 1350
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
1351 TATTAACACCAAAAAATAGCAAGAGGCTTAATATCCAGCTTAAGAAGG 1400
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
1401 TACAGAAAGTTTGGATTGAGCATCACTTCCAGAGATTAACAATTAAGTG 1450
484 lYserAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
1451 GCTCAGCTCCAAATCTATGTGAAAAACATTTCTCCCGGGGGGCGGCATT 1500
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGlnValAsnGlyVa 517
1501 CAGATGGCCGACTTAAGCAGAGACAGACTTATAGAGTAATGAGAGT 1550
517 laAspLeuValGlyLysSerGlnGluGlnValAlaSerLeuLeuArgSer 534
1551 AGATTTAGTGGCAAAATCCCAAGAGGAAGTTGTTCCGTGTGGAAGCA 1600
534 hLysMetGluGlyThrValSerLeuLeuValPheArgGlnLysAspAla 550
1601 CCAAGATGGAAGAACTGTGAGCTTCTGTGCTTTCGCCAGGAAGAGGCC 1650
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
1651 TTCACCCCAAGGAATGAAATGCAGAGCAAGCCAGATGCGAGATCCAAA 1700
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584
1701 AGAAACGAAAGCAGAAAGATGAGATTTGTTTACACTGATGGCACCA 1750
584 rGluPheLeuThrPheGluValAlaProLeuAsnAspSerGlySerAlaGly 600
1751 GGGATTTTCGACATTTGAAGTCCACTTAATGATTCAGATTCGACAGCC 1800
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
1801 CTGGTGTCAGTGTCAAAAGGTAAACGGGTCAAAAGAGAACACGACGATTT 1850
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
1851 GGGATCTTTGTCAAGTCAATTATTAATGAGAGACAGACATCTAAAGATG 1900

```

634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyIleSerLeu 650  
 |||||  
 1901 GAAGGCTTCGGGTGAATGATCACTGATAGCAGTAATGAGATATCCCTG 1950  
 |||||  
 651 LeuGlyIleThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667  
 |||||  
 1951 TTGGGCAAGACAAACCAAGATGCTCATGGAACCTTAAGAAAGCTTATGTC 2000  
 |||||  
 667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684  
 |||||  
 2001 TACTGAGGCAATTAACGAGGATGATCCAGCTATGCTTCAGAGGAA 2050  
 |||||  
 684 IeSerLysCysAsnGluLeuLysSerProGlySerProGlyProGlu 700  
 |||||  
 2051 TAAGCAAGTCCAAATGAGCTGAAGTCACTGAGGAGCCCTGAGACCTGAG 2100  
 |||||  
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717  
 |||||  
 2101 CTGCCCAATGAAACAGCGTTGGATGATAGAAAGAAATTTCCCATTC 2150  
 |||||  
 717 rLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734  
 |||||  
 2151 CCTTACGCTGGGATAGGGGCTTGATGAATCCCGACAGAAATGCTG 2200  
 |||||  
 734 IalSerArgIleMetGlyIleSerGlyLysTyrGlnLeuSerProThr 750  
 |||||  
 2201 CCTCAGTAGGATATG.....GCTAAATACCACTGCTCCCTAC 2241  
 |||||  
 751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767  
 |||||  
 2242 GTGAATATCCCAAGATGACACCTGCTATAGAAATGACAGGTGCC 2291  
 |||||  
 767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAsp 784  
 |||||  
 2292 AGTGCTCCCTCCACATCTCTGACAGTCCCTTCCAGCTCCCATGATG 2341  
 |||||  
 784 sPValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIle 800  
 |||||  
 2342 ATGTGGGGTTGTGACGCGCAGATGCTGTAAGTGGGCCCAAGCTCCAAATC 2391  
 |||||  
 801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817  
 |||||  
 2392 AGTGATTCAGCGCAGCTGCTTGAAGTCCAGATGTTGATCCAGTTCTTC 2441  
 |||||  
 817 aPheGluArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834  
 |||||  
 2442 TTTTCAACGAGAAAGATTGGAGCTCAGAGTATGTCAGAAAAACGACAA 2491  
 |||||  
 834 ySgLnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850  
 |||||  
 2492 ACCCAATTTCAGATGCCAGTCAATTGGATTTCCTTAACACCAAGAAATCA 2541  
 |||||  
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867  
 |||||  
 2542 AAAAGATGATTAGTATAGTATAGTACGACGACTAAACTCAATTAAGAGGA 2591  
 |||||  
 867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884  
 |||||  
 2592 TGACCCAGAAAGCAGGTTCTCCAGCAGAGAGATGSGGCTTCCTCCGGGTC 2641  
 |||||  
 884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900  
 |||||  
 2642 TGAAAGAGTCAAGCTCGTTGAGAGCTTCAGACCGCAGAGTTGCCAGGTG 2691  
 |||||  
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleLeuArg 917  
 |||||  
 2692 ACTTGTGATGGGATATCTTCCATCGTCCACGCGCGGATATATCAG 2741  
 |||||  
 917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyrA 934  
 |||||  
 2742 AGCGAGGGGATGCAATGAGAGCTTCAGAGCTGCGCATCGAACAAATCTTATG 2791

934 sPLysProAlaValAspAspAspAspGluGlyMetGluThrLeuGluGlu 950  
 |||||  
 2792 ATAAACCCCGCGGTAGATGATGATGAGAGCATGAGACCTTGGAAGAA 2841  
 |||||  
 951 AspThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlaSe 967  
 |||||  
 2842 GACACAGAAAGAAAGTTCAAGATCAGGAGAGAGAGCTGTATCCACAGCCAG 2891  
 |||||  
 967 rAspGlnProSerHisSerLeuGluArgGlnMetAsnGlnLysAsnGlnL 984  
 |||||  
 2892 TGATCAGGCTTCCACTCTCTGAGAGACAAATGATGGAACCAAGAGGA 2941  
 |||||  
 984 ySgGlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLys 1000  
 |||||  
 2942 AAGGTGATTAACCTGATAGAAAAAGATTAACCTGAAAAAGAAAGAG 2991  
 |||||  
 1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017  
 |||||  
 2992 AAGATAGAGATTAAGAGAAAGATTAATTAAGCAAGAAAGGAATGCT 3041  
 |||||  
 1017 uLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAspL 1034  
 |||||  
 3042 GAGGCGCTTGGGAGACATGTTGAGTTGGCAACATCGAAAGATGACA 3091  
 |||||  
 1034 yAlIleGlyLysThrGlyLysIleLysIleGlnGluSerPheThrSerGlu 1050  
 |||||  
 3092 AGATTGAGAAACGGGTAAATTAATTAACAGATCCTTATACATCAGAA 3141  
 |||||  
 1051 GluGluArgIleArgMetLysGlnGluGlnGluArgIleGlnAlaLysThr 1067  
 |||||  
 3142 GAGAGAGAGATTAAGAAATGAAGCGAGAGAGAGATTCAGCAAGAAC 3191  
 |||||  
 1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyrAlaGluIleG 1084  
 |||||  
 3192 TCGAATATTAGGAGACACACAGCTCGAAGCGCTGACTGCTGAATTC 3241  
 |||||  
 1084 IAspPheHisArgThrPheGlyCysAspAspGluLeuMetTyrGlyGly 1100  
 |||||  
 3242 AAGATTTTCATCGGACATTGGCTGTGATGATGATGATTAATGTTGGGGA 3291  
 |||||  
 1101 ValSerSerTyrGluGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117  
 |||||  
 3292 GTTCTCTTATGAAGGTTCCATGGCTCTCAACGCTAGACCTCAGAGCC 3341  
 |||||  
 1117 oArgGluGlyHisMetMetAspAlaLeuTyrAlaGlnValLysLysPro 1134  
 |||||  
 3342 ACGAGAAAGGCATATGATGATGATGCTTGTATGCCCAAGTCAAGAGCGC 3391  
 |||||  
 1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150  
 |||||  
 3392 GGAATTCCAAAACCTCCTGCTAGACAGTAACAGATCAACTCTTAGCAAT 3441  
 |||||  
 1151 HisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAs 1167  
 |||||  
 3442 CATGATCGATACAGCGTCTGAGGCAGAAATTTCAAGCAAGCAAGCAAGA 3491  
 |||||  
 1167 pGluAspValGluAspArgArgArgThrTyrSerPheGlnGlnProCTPP 1184  
 |||||  
 3492 TGAAGATGTAGAAAGATGCTGCGCGGACCTATAGTTTGAGCAACCTGGC 3541  
 |||||  
 1184 rOAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200  
 |||||  
 3542 CGAAGCCAGCGCGCGCAGCGAGCGGCGGCGACACTCGGTGCGCGAG 3591  
 |||||  
 1201 ValGlnMetGlnArgGlnArgGlnGluGluGluLysSerLeuGlnL 1217  
 |||||  
 3592 GTGCAATGCAAGCGGCAAGCGCAGAGAGCGCAGAGCTCCAGAGGC 3641  
 |||||  
 1217 aGlnArgGlnTyrSerSerLeuProArgGlnSerArgLysAsnAlaSer 1234  
 |||||  
 3642 CCAGCGCCAGTACAGCTCTGCTCGGCAAGCAAGAAAAATGCCAGCT 3691  
 |||||  
 1234 eValSerGlnAspSerTrpGluGlnAsnTyrSerProGlyGluGlyPhe 1250

|||||  
3692 CGGTCTCCAGGACTTGGAGCAGAACTACCTCCCTGGGGAAGGCTTC 3741  
1251 GlnSerAlaIysGlnAspProArgTyrSerSerTyrGlnIysSerArgas 1267  
3742 CAAAGTCCCAAGAAAGAACCCAGTACTCCAGTACCAAGGCTCCAGAA 3791  
1267 nglYtYrLeuGlyGlyHISglYpheaSnaIaArgValMetLeuGlnUthrG 1284  
3792 CGGCTACCTGGAGACATGCTTCAACGCGCAGGATCATCTCTGAAATC 3841  
1284 InLluLeuLeuArgInGlnGlnArgArgYsgIuGlnGlnMetIysLys 1300  
3842 AGGAGCTCTCCGCGCAGAACAGAGCGGAGAGCAGCAGATGMAAG 3891  
1301 GlnProSerGlnIysProSerAsnTyrAspSerTyrIysLysValG 1317  
3892 CAGCTTCCTCCAGAGGCGCCAGCACTAGCTCGATTAAGAAATGCCA 3941  
1317 naSPProSerTyrAlaProIProIysGlyProPheArgGlnAspValProp 1334  
3942 GGAGCCAGATTACGCCCCCTCCCAAGGGGCGCTCCGCGCAGATGTGCC 3991  
1334 roSerProSerGlnValAlaArgLeuSnaArgLeuGlnThrProGlnLys 1350  
3992 CCTCCCTCTCAGGTGCGAGCTGAAACAGACTTCAGACTCTCGAGAAA 4041  
1351 G1ArgProPheTyrSer 1356  
4042 GGGAGGCGCTTCATTC 4059

seq\_name: gp\_pr:AF196185

seq\_documentation\_block:

LOCUS AF196185

DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein long

ACCESSION AF196185

VERSION AF196185.1

KEYWORDS

SOURCE

ORGANISM

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS Fang, C.M. and Xu, Y.H.

TITLE Down-regulated expression of atypical PKC-binding domain deleted

JOURNAL Cell Res. 11 (3), 223-229 (2001)

MEDLINE 2152080

PUBMED 11642408

REFERENCE  
AUTHORS Fang, C. and Xu, Y.

TITLE Exon/Intron Structure and Splicing Variants of a Novel Human

JOURNAL Unpublished

REFERENCE  
AUTHORS Fang, C. and Xu, Y.

TITLE Direct Submission

JOURNAL Submitted (18-OCT-1999) Laboratory of Molecular and Cellular

FEATURES  
source

1. .5958

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="10"

/map="10p11.2"

288. .4349

/note="hasIP1a"

/codon\_start=1

/product="atypical PKC isotype-specific interacting

protein long variant"

/protein\_id="AAK27891.1"  
/db\_xref="GI:13491610"  
/translation="MKVYVCFGRTRRVVPCGDGHMKVFSLIQOAVTRYKALAKDPNY  
WIOVRLEHGGGIIIDLDIIICDVADDDKRLAVLVEEDDPHHGGGTSSNSGQSP  
IIGSRIGTNNVSAPOPCATSEIYTPVSLRANLHVRRSSDPALIGISTVSNSP  
SSEEPSRKPTRWSTTAGTFLKNTGASPTCDKRDENRSLPRTSMNSQFQDNA  
RSSLSASHPHWKLEKQEDGTEEDNSREVPVHAQTGLEHLPNESLDMKLYE  
VNDGPPDIIHVPPSARIGTGLVYKLEKLEAEHNLRENDCLYRINDMGLRN  
RRFEQOAHFRQAMRTPILMFVVAANKEOYEOSREKNYYSRSFSDQYIDNR  
SVNSAGLHVQAPRLNHPPEQIDSHSRLPHSAHPGKPPSPAPASPOVFSTVSSG  
YVTKIKGRNLNOLKCKTGGLFSTSDPTGCSAPITVKNILPRGAILDGRILKAG  
DRLIEVNGVDLYGKSOEEVSLRSTKMEGTSLVPROEDAFHRELAEPSQMDIP  
KETKADEDIVLTPDGTREFLEVPALNDGSAAGLYSVKGRSKENHADLGI1FVKSI  
INGGASAKGRRLVNDOLIAVNGESILGRTNDAAETLRSMSTEGNKGM1QLIVAR  
RISKNEILKSPSPGPELPETALDDRETRISLSISGIEGLDPSRNALSTRMG  
KYOLSPYVMPPQDDVYLIEDDLRPLVPLPHLSOSSSSSSSSSSSSSSSSSSSS  
DSADSLSTDPVYVLAFOREGGRGROSMSEKPRKQSDASQDLFPVTRKSKMDLIAD  
ETKLTIVDQKAGSPSRDVGPSLGILKSSLSLQTVAVELVNLNDIPEHRRPRLIR  
GRGNESEFRALADKSYDKPAVDDEGMEETLEDETESSRSRESVSTASDPSHLE  
ROMNNOEKGDTRDKKDKTEKKRDKRDKKAKKAKKGLGDMRFGRKHRRKD  
KLEKTGIRKIOEFTSEERIRMKOQOERIOAKTRFERROAREYDARIDOFHRTG  
CDEELMVGSSVEGSMALNARPOSPREGHMDALYAOVKKPRNSKPSVDVNSRTPS  
NHDRIORLROEFOAKODPDVDRRRTYSFQDPMPARATQSGHSVSVEYOMKOR  
QKERSSQOAGQOYISLPSQSKNMSVSQDSMEDONTSGEFGSKENRPTSS1QGS  
RNGYLAGEHFNARVMLETQELRLROERREDDQMKRQPPSEGSVNSDYKRYDDPSYAP  
PKGPRQDPSPSOVARLNLRLQTEPEKGRPTFS"

BASE COUNT 1623 a 1395 c 1581 g 1358 t 1 others

alignment\_scores:

Quality: 7008.50

Ratio: 5.180

Percent Similarity: 99.779

Percent Identity: 99.779

alignment\_block:

us-09-757-781-2 x AF196185 ..

Align seg 1/1 to: AF196185 from: 1 to: 5958

1 MetLysValThrValCysPheGlyArgThrArgValValProGlyG 17  
|||||  
288 ATGAAAGTGAACCGTGTCTGGAGGACCGGGGTGGTGGCGTGGGG 337  
17 yASpGLHISMetLysValPheSerLeuIleGlnAlaValThrArgT 34  
338 GGAGGCGCACATGAAGTTTTCACCTTCATCCACAGCGGTGACCCCT 387  
34 yTArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50  
388 ACCGGAAGCCCATGCCAAGGATCCAAACTGATGATACAGTGCATGCC 437  
51 LeuGlnHisGlyAspGlyIleLeuAspLeuAspAspIleLeuGlyAs 67  
438 TTGGAACATGGAGATGGAGAACTAGACTTGATGACATCTTGTGGA 487  
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnAspP 84  
488 TGTAGCAGACATGAAGACAGACAGTACGCTTTTGATGACGAGATC 537  
84 roHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100  
538 CACATTCACGAGGATGATGCCACAGTCCAGTCCACGCGGATCCAGAGC 587  
101 ProGlnIlePheGlySerGlnLeuGlyThrTrpAsnAsnValSerAlaPheG 117  
588 CCAAGATATTGTAGTGGAGCTTGCGCACCAATGCTCTAGGCTTTCA 637  
117 nProTyrGlnAlaThrSerGlnIleGlnValThrProSerValLeuArg 134  
638 GCGTTACCAAGCAACAGTGAATTTGAGTGCACACCTTCATCTTGGAG 687  
134 laAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150

```

|||||
688 CAAATATGCTCTTCAATGTTCCAGCAGATGACCCAGCTCTAATGTC 737
151 LeuSerThrSerValSerAspSerAnpSerSergLugLupProSerAr 167
|||||
738 CTCTCCACTTCTGTGATGATATTTTCTCTGGAAGACCTTCAAG 787
167 glysasnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184
788 GAAAAATCCACACGCTGGTCAACAACACCTGGCTTCTCAAGCAGAAACA 837
184 hrAlaGlySerProLysThrCysAspArgLysAspGluAsnTyraG 200
838 CTGCTGGAGATCTCTAAACCTGGACAGAGAAAGATGAAACCTACAGA 887
201 SerLeuProArgAspThrSerAnpTrpSerAnpGlnPheGlnArgAspAs 217
888 AGCCTCCCGGGGATCTAGTAACTGTTCAACCAATTTTCAGAGACAA 937
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
938 TGGTCCGCTCTCTGATGTCAGTCCAGTACCCAAATGGTGGCAAGTGGCTGG 987
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
988 AGAAACAAAGACAGATGAGATGGGACAGAGAGATAAACAGTGTGTT 1037
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPhese 267
1038 GAAACCTGGTGGACATGCTGCACAGGGTTGGAGCATATACCCAACTTTTC 1087
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysIProL 284
1088 TCTGGATGATATGCTAAAGCTCTCTAGAACTCCCAACGATGGAGGGCTTC 1137
284 euGlyIleHisValIleProPheSerAlaArgGlyLysArgThrLeuGly 300
1138 TGGGAATCCATGTAGTGCCTTTCAGTGCCTCGAGGCGGACAGAACCTGGGG 1187
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisGluAsnIle 317
1188 TTATTAGTAAACGATTGGAAAGGTGTGAAGCTGAACATATAAATCT 1237
317 upheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
1238 TTTTCTGGAATATGATTCATTTCTCAGGATTAATGATGGGACCTTCGAA 1287
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
1288 ATAGAGATTTGAAACAAGCACAACATATGTTTCCCAAGCCATGCGTACA 1337
351 ProIleIleTrpPheHisValIleProAlaAlaAsnLysGluGlnTyraG 367
1338 CCCATCATTTGGTCCATGTGTTCCGCGACAATATGAAGACAGATATGA 1387
367 uGlnLeuSerGlnSerGluLysAsnAsnTyraTyraSerSerArgPheSerP 384
1388 ACAACTATCCCAAGTGAGAAACAATTACTATTCAGCCGTTTATGCC 1437
384 roAspSerGlnTyraIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
1438 CTGACAGCCAGTATATTGACMACAGAGTGTGAACAGTGCAGGGCTTCAC 1487
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
1488 ACGGTGAGAGAGACCCGAGCTGAACACCCGCTGAGCAGATAGACTC 1537
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
1538 TCACTCAAGACTACCTCATAGCGACACCCCTCGGGAACCAACCAATCCG 1587
434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSergLy 450
|||||

1588 CTCCAGCCTCGGACACCTCAGATGATATTAGTACGACTGTAAAGCAGTGT 1637
451 TyraAsnThrLysLysIleGlyLysArgLeuAsnIleGlnIleLysLysG 467
1638 TATAACACCAAAAAATAGGCAAGAGGCTTAATATACAGCTTAAGAAAGG 1687
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484
1688 TACAGAGAGTTTGGGATTCAGCATCATTCCAGAGATGATGAACATATAGGG 1737
484 LySerAlaProIleTyraValLysAsnIleLeuProArgGlyAlaAlaIle 500
1738 GCTCAGCTTCCAACTATGTAAGAAAAACATTTCCCGGGGGGGCGGCATTT 1787
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517
1788 CAGGATGGCCGACTTAAGCAGAGACACAGACTTATAGAGTAAATGCAAT 1837
517 laSpleuValGlyLysSergLugLugLuvAlaValSerLeuLeuArgSert 534
1838 AGATTTAGTGGCAATCCCAAGAGGAAGTGTTCGCTGTGAGAGACGA 1887
534 hrLysMetGluGlyThrValSerLeuLeuValPheArgGlnLysPAla 550
1888 CCAGATGGAGAGAACTGTGAGCTTGTGCTTCCGCGCAGGAAGACGCC 1937
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
1938 TTCCACCCCAAGGAACTGAATCCAGACGCCAACCAGATCCAAATTCCAA 1987
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrA 584
1988 AGAAACGAAGCAAGAAAGATGAGATATTTGTTCTTACCTCAGTGCACACA 2037
584 rGluLupheLeuThrPheGluValProLeuAsnAspSerLysSerIleGly 600
2038 GGAATTTCTGACATTTGAAGTCCCACTTAATGATTCAGAGATCTCCAGGC 2087
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspIle 617
2088 CTTTGGTGTACGTCTCAAGTAAACCGTCAAAAAGAAACACGACGATTTT 2137
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
2138 GGAATCTTTGTCAAGTCCATTAATGAGAGAGACGATCTAAAGANG 2187
634 LyArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerLeu 650
2188 GAGGCTTCGGGTGAATGATCAACTGATAGCAGTAATGAGAAATCCCTG 2237
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
2238 TTGGGCAAGCAAAACCAAGTGGCATGGAACCCCTAAAGAGTCTATGTC 2287
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684
2288 TACTGAAGCAATAAACGAGGAATGATCCAGCTTAATGTTGCAAGGAGAA 2337
684 leSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700
2338 TAAGCAAGTGCATATGAGTGAAGTCACTGGAGAGCCCTTGACACTGAG 2387
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
2388 CTGCCCATTTGAACACAGCTTGGATGATAGAGAAAGAAATTTCCCATTC 2437
717 rLeuTyraSergLyIleGluGlyLeuAspGluSerProSerArgAsnAlaA 734
2438 CCTCTACAGTGGGATTTGAGGGCTTGAATGAAATGCCACAGAAATGCTG 2487
734 laLeuSerArgIleMetGlyLysSergLyLysTyraGlnLeuSerProThr 750
2488 CCTCTCAGTAGATTAATG.....GTTAAATACCAAGCTGTCCCTTACA 2528

```



751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767  
|||||  
2529 GTGAATATGCCCCAAGATGACACTGTCAATATAGAGATGACAGGTGCC 2578  
767 OValLeuProProHISLeuSerAspGlnSerSerSerSerHisAspA 784  
2579 AGTGTCTCTCCACATCTCTGTACACAGTCTCTTCCAGCTCCCAATGATG 2628  
784 sPValGlyPheValThrAlaAspAlaGlyThrTrpAlaIleIle 800  
2629 ATGTGGGGTTTGTGACGACGATGCTGTACTTGGGCCAAGCGTCGAATC 2678  
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817  
2679 AGTGATTACAGCCGACTGCTCTTGGAGTCCAGATGTTGATCCAGTCTGTGC 2728  
817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834  
2729 TTTTCAACAGAAAGATTGGACGTGAGATGTCAGAAAAAGCACA 2778  
834 ySGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850  
2779 AGCAATTTTCAGATGCCAGTCATTTGGATTTCCTTAAACACGAATATCA 2828  
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867  
2829 AAAACGATGATTAGTATAGCTGACGAGACTAAACTCAATACAGTGA 2878  
867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884  
2879 TGACCGAAAGCAGGTTCTCCAGCAGAGATGGGTCTCTCCCTGGGTC 2928  
884 eulysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900  
2929 TGAAGAAGTCAAGCTCATTTGAGAGTCTGCAGACCCAGTTGCCAGGTC 2978  
901 ThrLeuAsnGlyAspIleProPheHISArgProArgProArgIleIleArg 917  
2979 ACTTTGAATGGGGATTTCTCTTCCATCGTCCACGGCCGGGATTAATCAG 3028  
917 GGIArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerThrA 934  
3029 AGGCAGGGATGCATGAGAGCTTCAGAGTGCATCGACAAATCTTATG 3078  
934 sPlysProAlaValAspAspAspAspGluGlyMetGluThrLeuGlnGlu 950  
3079 ATTAACCCCGGTAGATGATGATGAAGCGCATGGAGACCTTGGAAGAA 3128  
951 AspThrGluGlnLysSerArgSerGlyArgGluSerValSerThrAlaSe 967  
3129 GACACAGAAAGTAAGTCAAGATCAGGAGAGAGTCTGTATCCACACAGCAG 3178  
967 rAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlyAsnGlnGluL 984  
3179 TGTATCAGCCCTCCACTCTCTGTGAGAGACAAATGAATGGAACCAAGAGA 3228  
984 ySGlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000  
3229 AAGGTATTAAGACTGATAGAAAAAAGATATAAACTGAAAAAAGAAAGAAC 3278  
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017  
3279 AAAGATAGAGATTAAGAGAAAGATTAATAATGNAAGCAAGAAAGGAATGCT 3328  
1017 uLysGlyLeuGlyLysMetPheArgPheGlyLysHisArgLysAspAspL 1034  
3329 GAAGGGCTTTGGGACATGTTTCAGTTTGGCAAAACATCGAAAAAGATGACA 3378  
1034 ysiIleGlyLysThrGlyLysIleLysIleGlnGlnLysPheThrSerGlu 1050  
3379 AGATTGAGAAAAAGGGTAAAAATAATACAGGAATCCTTTACATCAGAA 3428

1051 GluGluArgIleArgMetLysGlnGlnGlnGluArgIleGlnAlaLysTh 1067  
|||||  
3429 GAGGAGAGGATACGAATTAAGACAGAGACAGAGAGATTCACAAAC 3478  
1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspArgIleGluIleG 1084  
3479 TCGAAGATTTTAGGGAACGACAAAGCTCAGAGCGTGACTATGTGAATTC 3528  
1084 InAspPheHisArgThrPheGlyCysAspAspGluLeuMetLysGlyGly 1100  
3529 AAGATTTTCAATCGACATTTGGGCTGTGATGATGATTAAGTATGGGGA 3578  
1101 ValSerSerTyroGluGlySerMetAlaLeuAsnAlaArgProGlnSerP 1117  
3579 GTTCTCTTATGAAGGTTCCATGGCTCTCAACGCTAGACCTCAGAGCC 3628  
1117 oArgGluGlyHisMetMetAspAlaLeuTrpAlaGlnValLysLysPro 1134  
3629 ACGAGAAAGGCATATGATGATGATGCTTTGTATGCCAAAGTCAAGAAAGCCG 3678  
1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150  
3679 GGAATTCGAACCCCTCACCTGTAGACAGTAAACAGATCAACTCCTAGCAAT 3728  
1151 HisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAs 1167  
3729 CATGATTCGATACAGCGCTCTGAGGCAAGAAATTCACGACAAAGCAAGA 3778  
1167 pGluAspValGluAspArgArgArgThrTyroSerPheGluGlnProTrp 1184  
3779 TGAAGATGTAGAAAGATCTGTGGCGAGACTTATAGTTTGACCAACCCGTGC 3828  
1184 rOAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200  
3829 CGAAGCAGAGCGCGGCGAGCAGCAGAGCGGCGACACTGGGTGCTGTGAG 3878  
1201 ValGlnMetGlnArgGlnArgGlnGluGluArgGluSerSerGlnGlnAl 1217  
3879 GTGCAGATGCAGCGCGAGCCCGAGAGAGAGCGGAGAGACTCCACAGCGGC 3928  
1217 aGlnArgGlnTyroSerSerLeuProArgGlnSerArgLysAsnAlaSerS 1234  
3929 CCAAGCGCATACAGCTCTGTGCTCGGCAAGCAGGAAAAATGCCAGCT 3978  
1234 eValSerGlnAspSerTrpGluGlnAsnTyroSerProGlyGluGlyPhe 1250  
3979 CGGTCTCCAGGACTCTTGGAGCAGAACTACTCCCTGGGGAAGGCTTC 4028  
1251 GlnSerAlaLysGluAsnProArgTyroSerSerTyroGlnLysSerArgS 1267  
4029 CAGAGTGCCAAAAGAAAGCCAGGTACTCCAGCTACCAAGAGCTCCAGAA 4078  
1267 nGlyTyroLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGluThrG 1284  
4079 CGGCTACTGTGAGAGATGCTTCAAGCGCAGGCTATGCTGTGAAACTC 4128  
1284 InGluLeuLeuArgGlnGlnGlnArgArgLysGlnGlnGlnMetLysLys 1300  
4129 AGGAGCTCCTTCGCCAGAAACAGAGGCGGAAGAGAGCAGCATGTAAGAAAG 4178  
1301 GlnProProSerGluGlyProSerAsnTyroAspSerTyroLysLysValG 1317  
4179 CAGCTCTCTTCCAGAGGGGCCACAGCACTATAGACTCGATTAAGAAAGTCCA 4228  
1317 nAspProSerTyroAlaProProLysGlyProPheArgGlnAspValaPro 1334  
4229 GGAAGCCAGTTACGCCCTCCCAAGGGGGCCCTTCGGCAAGATGTGGCCC 4278  
1334 rOserProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGluLys 1350  
4279 CTTCCCTTCTCAGGTTCGAGGCTGAAACAGACTCTGAGACTCTCTGAGAA 4328  
1351 GlyArgProPheTyroSer 1356



701 AGAACACACAGATGAGATGGACACAGAGATTAACAGTCCGTGTT 750  
 251 GIUProValGIyHISAlaasprhrgiLeuGluHISileProaspHse 267  
 751 GAACCTGTTGACATGCTGACAGCGGTTTGAGCATATATACCAACTTTTC 800  
 267 rLeuaspaspmetValLysLeuValGIuValProaspaspGlyLysProl 284  
 801 TCTGGATGATATGGTAAAGCTCTAGACGCCCAACGATGGAGGCGCTC 850  
 284 euGIyLLeHISValIlePropheserAlaargGIyGIyArgrHLeuGIy 300  
 851 TGGGAATCCATGATGATGCTTTCACGTCTGAGCGCGCAGAACCTTGGG 900  
 301 LeuLeuValLysArgrLeuGIuLysGIyGIyLysAlaGIuHISGluasnl 317  
 901 TTATTTAGTAAACGATTTGGAAAGGTGGTAAAGCTGAACATGAAATCT 950  
 317 uPheArgrGIuaspaspCysGIeValArgrIleaspaspGlyAspLeuArgr 334  
 951 TTTTTCGTGAAATGATTCATGCTGAGGATTAATGATGGCGACCTTCGAA 1000  
 334 snArgrArgrPheGIuGIuAlaGIuHISmerPheArgrGIuAlaMetArgrThr 350  
 1001 ATAGAAAGATTGAAACAGCACACATATGTTTCCCAAGCCATGCGTACA 1050  
 351 ProIleIleTrPheHISValValProAlaAlaAsnLysGIuInTrgrI 367  
 1051 CCCATCATTTGGTTCCATGTGGTCCGACGCAAAATAAAGACGATATGA 1100  
 367 uGIuLeuSerGIuInSerGIuLysaspasnlTrYTrSerSerArgrPheSer 384  
 1101 ACACATATCCCAAGAGAGAAACAATTAATCAATTCAGGCCGTTTAAACC 1150  
 384 roaspserGIuInTrIleaspaspArgrSerValaspSerAlaGIyLeuHIS 400  
 1151 CTACAGGCCATGATATTGACACACAGAGTGTGAACAGTGCAGGCGTTTAC 1200  
 401 ThrValGIuArgrAlaProArgrLeuaspasnlProGIuGIuInIleaspse 417  
 1201 ACGGTGACAGAGACACCCGACGTGAACACCCCGCTGAGCAGATAGACTC 1250  
 417 rHISerArgrLeuProHISerAlaHISProserGIyLysProProserA 434  
 1251 TCCACTCAAGCTACATACATAGCCACACCCCTGGGAAACCAACATCCG 1300  
 434 laProAlaSerAlaProGIuAsnValPheSerThrThrValSerSerGIy 450  
 1301 CTCACAGCTCGCACCTCAGAAATGATTATTAGTACGACTGTAAAGCATGGT 1350  
 451 TrYAsnThrLysLysIleGIyLysArgrLeuAsnIleGIuLeuLysLysGI 467  
 1351 TATTAACCCCAAAAATAATAGCAGAGAGGCTTAATATCCAGCTTAAGAAAG 1400  
 467 YThrGIuGIyLeuGIyPheSerIleThrSerArgrAspValThrIleGIy 484  
 1401 TACAGAGAGGTTTGGATTCAGCATCACTTCACAGATGTAAACAATAGCTG 1450  
 484 LysSerAlaProIleTrYValLysAsnIleLeuProArgrGIyAlaAlaIle 500  
 1451 GCTCAGCTCCAAATCTATGTGAAAAACATTTCTCCCGGGGGGGGCGCATTT 1500  
 501 GluAspGIyArgrLeuLysAlaGIyAspArgrLeuIleGIuValaspnGIyVal 517  
 1501 CAGGATGGCGCGACTTAAGCAGAGAGACAGACTTATAGAGGTAATGTAGT 1550  
 517 laAspLeuValGIyLysSerGIuGIuValValSerLeuLeuArgrSerT 534  
 1551 AGATTAGTGGCAAAATCCCAAGAGAAAGTGTTCGCTTGGAAGACA 1600  
 534 hrYlYsMetGIuGIyThrValSerLeuLeuValPheArgrGIuGIuLysAla 550  
 1601 CCAAGATGGAAGAACTGTGAGCTTCTGCTTTTCCGCAAGAAAGACGCC 1650

551 PheHISProArgrGIuLeuaspasnlAgIuProSerGIuMetGIuInIleProLy 567  
 1651 TTCCACCAAGGAACTGATTCAGAGACCACCAAGCCAGATGCAGATTCCAAA 1700  
 567 sGIuThrLysAlaGIuAspGIuAspIleValIleLeuThrProaspGIyThra 584  
 1701 AGAAGCGAAAGCAGAAAGATGAGATATGTTGTTTACACCTGATGGCCACA 1750  
 584 rGSIuPheLeuThrPheGIuValProLeuaspasnsPserGIySerAlaGIy 600  
 1751 GGGAAATTCGACATTTGAAAGTCCCACTTAATGATTTAGGATGTGCGAGC 1800  
 601 LeuGIyValSerValLysGIyaspasnlSerLysGIuaspasnlAlaAspIle 617  
 1801 CTTGGTGTCAGTGCAAAAGGTAAACCGGTCAAAAGAGAACACCGAGATTT 1850  
 617 uGIyIlePheValLysSerIleIleaspnGIyGIyAlaAlaSerLysasp 634  
 1851 GGGAAATCTTGTCAAGTCCATTTATTAATGAGAGAGAGCATCTAAAGATG 1900  
 634 LysArgrLeuArgrValaspasnlLeuIleAlaValaspnGIyGIuSerLeu 650  
 1901 GAAGGCTTCGGGTGAATGATCACTGATAGCAGTAAATGGAGAAATCCCTG 1950  
 651 LeuGIyLysThrAsnGIuaspAlaMetGIuThrLeuArgrArgrSerMetSe 667  
 1951 TTGGGCAAGACAAACCAAGATGCCATGGAACCTTAAGAGGTCTATGTC 2000  
 667 rThrGIuGIyAsnLysArgrGIyMetIleGIuLeuIleValAlaArgrArgr 684  
 2001 TACTGAAGGCAATTAACGAGAAATGATCCAGCTTAATGTTGTCAGAGGAGA 2050  
 684 leSerLysCysAsnGIuLeuLysSerProGIySerProGIyProGIu 700  
 2051 TAAACAGTGCATGATGACTGAAGTCACTGGAGACCCCTTGAGACTGAG 2100  
 701 LeuProIleGIuThrAlaLeuaspasnlArgrGIyIleSerHISse 717  
 2101 CTGGCCATTGAACAGCGGTTGATGATAGAGAAAGAAATTTTCCCATTC 2150  
 717 rLeuTrYSerGIyIleGIuGIyLeuaspGIuSerProSerArgraspAla 734  
 2151 CCTCTACAGTGGAGATTGAGGGGCTTGATGAATCCCGCAGCAAAATCTG 2200  
 734 laLeuSerArgrIleMetGIyGIuSerGIyLysTrGIuLeuSerProThr 750  
 2201 CCTCAGATGATATATGAGTGAAGTCAAGTAAATCCAGCTGTCCCTTACA 2250  
 751 ValaspasnlProGIuaspasnlThrValIleIleGIuaspasnlArgrLeuPr 767  
 2251 GTGAATATGCCCCCAAGATGACACTGTCATTATAGAAGTGAAGAGTTGCC 2300  
 767 oValLeuProProHISeraspasnlSerSerSerSerSerHISaspA 784  
 2301 AGTGCTTCCCTCCACATCTCTCTGACAGTCCCTTCCAGCTCCCATATG 2350  
 784 spValGIyPheValThrAlaaspAlaGIyThrTPAlaLysAlaAlaIle 800  
 2351 ATGTGGGTTTGTACCGGACGATCTGTACTTGGCCAMGGCTGCAATC 2400  
 801 SeraspSerAlaaspCysSerLeuSerProaspValaspProValLeuAl 817  
 2401 AGTATTTCAAGCGCAGCTGCTTTGAGTCCAGATGTGATTCAGATTCTTGC 2450  
 817 aPheGIuArgrGIuGIyPheGIyArgrGIuSerMetSerGIyLysArgrThrL 834  
 2451 TTTTCAACGAGAAAGATTTGACGTCAAGATATCTCAAGAAAAAGCACAA 2500  
 834 YsGIuPheSeraspAlaSerGIuLeuaspPheValLysThrArgrLysSer 850  
 2501 AGCAATTTTCAGATGCCAGTCAATTTGATTTCTTTAAACAGCAAAATCA 2550

851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867  
 2551 AAAACATGATTTAGATAGCTACAGCACTAACTCAATACAGTGA 2600  
 867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884  
 2601 TGACCGAAGAACAGGTCTCCACGACAGATGTGGTCTCCCTGGGTC 2650  
 884 eulLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900  
 2651 TGAAGAAGTCAGCTGTTGGAGAGTCGACAGCCGAGTTGCCGAGGTC 2700  
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgIleIleArg 917  
 2701 ACTTTGAATGGGATATCTCTTCATCTCCACGCGCGGATATATCAG 2750  
 917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerTyrA 934  
 2751 AGCGAGGGGATGCAATGAGAGCTTCAGAGCTGCCATCGACAAATCTTATG 2800  
 934 sPLysProAlaValAspAspAspAspGluGlyMetGluThrLeuGlnGlu 950  
 2801 ATAAACCCGCGGTGATGATGATGATGATGATGATGATGATGATGATGAT 2850  
 951 AspThrGluLysSerSerArgSerGlyArgGluSerValSerThrAlaSe 967  
 2851 GACACAGAAAGAGTTCAGATCAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2900  
 967 rAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlyAsnGlnGlu 984  
 2901 TGATCAGGCTCCCTCTCTGAGAGACAAATGATGATGATGATGATGATGAT 2950  
 984 yGlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000  
 2951 AAGGTGATAGACTCATAGAAAAAAGATAAACGTGAAAAAGAAAAAGAG 3000  
 1001 LysAspArgAspLysGlnLysAspLysMetLysAlaLysLysGlyMetLe 1017  
 3001 AAAGATAGAGATAGAGAGAGAGATAAATGAAAGCCAGAAAGGAGATCCT 3050  
 1017 uLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAspL 1034  
 3051 GAAAGGCTGGGAGACATCTTC..... 3072  
 1034 ySileGluLysThrGlyLysLleLysLleGlnGluSerPheThrSerGlu 1050  
 3072 ..... 3072  
 1051 GlnGluArgLleArgMetLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1067  
 3073 .....AGGATTCAGAGCCCAAAAC 3089  
 1067 rArgGluPheArgGlnArgGlnAlaArgGlnArgGlnArgGlnArgGlnG 1084  
 3090 TCGAAGATTTAGGAGACAGACAGAGCTGAGCGCTGACTAGCTGGAATTC 3139  
 1084 lAspPheHisArgThrPheGlyCysAspAspGlnLeuMetTyrGlyGly 1100  
 3140 AAGATTTCATCGACATTTGGCTGCTGATGATGATGATGATGATGATGATG 3189  
 1101 ValSerSerTyrGlnGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117  
 3190 GTTCTCTTATAGAGCTTCATGCTTCACCTCAGCTCAGAGCTCAGAGCCC 3239  
 1117 oArgGlnGlyHisMetMetAspAlaLeuTyrAlaGlnValLysLysProA 1134  
 3240 ACGAGAAAGGCGATATGATGATGATGATGATGATGATGATGATGATGATG 3289  
 1134 rGlnAsnLysLysProSerProValAspSerAsnArgSerThrProSerAsn 1150  
 3290 GGAATTCCAAACCCCTCAGCTGAGACAGTAAACAGATCAACCTCAGCAT 3339  
 1151 HisAspArgLleGlnArgLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGln 1167

3340 CATGATCGGATACAGCTCTGAGCCAGAAATTTACAGCAAGCAAGCAAGA 3389  
 1167 pGluAspValGlnAspArgArgArgArgArgArgArgArgArgArgArgArg 1184  
 3390 TGAGATGTAGAGATCGTGGCGGACCTATAGTTTGAAGCAACCTGGG 3439  
 1184 rAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200  
 3440 CGAAGCCAGCGCGCGAGCGAGCCGAGCCGAGCCGAGCCGAGCCGAGCCG 3489  
 1201 ValGlnMetGlnArgGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnG 1217  
 3490 GTTCAGATTCAGAGCGGAGCGGAGGAGGAGCGGAGGAGCTCCAGCAGGC 3539  
 1217 agLinarGlnTyrSerSerSerLeuProArgGlnSerArgLysAsnAlaSer 1234  
 3540 CCAGCCGACGTACAGCTCTCTCTGCGCAAGCAAGCAAGCAAGCAAGCT 3589  
 1234 eValSerGlnAspSerTrpGlnGlnAsnTyrSerProGlyGluGlyPhe 1250  
 3590 CGGCTCTCCAGGACTCTTGGAGAGAACTACTCCCTGGGAGAGGCTTC 3639  
 1251 GlnSerAlaLysGlnAsnProArgTyrSerSerTyrGlnGlySerArgAs 1267  
 3640 CAGAGTCCCAAGAGAAAGCCAGAGTCTCAGCTACCAAGGCTCCAGGAA 3689  
 1267 nGlyTyrLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGluThrG 1284  
 3690 CGGCTACTCGGAGGAGATGGCTTCAACGCGCAGGCTCATCTGGAAATCTC 3739  
 1284 lngLleuLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 1300  
 3740 AGAGCTCTCTCCGAGGAAACAGGCGGAGGCGGAGGCGGAGGCGGAGGAG 3789  
 1301 GlnProProSerGlnGlyProSerAsnTyrAspSerTyrLysValGln 1317  
 3790 CAGCTCTCTCCGAGGCGGCGGAGCACTATGACTCGTATAGAAAGTCCA 3839  
 1317 nAspProSerTyrAlaProPolysGlyProPheArgGlnAspValProp 1334  
 3840 GGACCCCGATTACGCCCTCCCAAGGGGCGCTTCGGCAAGATGGCCCC 3889  
 1334 rSerProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGlnLys 1350  
 3890 CCGCCCTCTCAGGCTCGAGGCGAGGCGAGACAGCTTCAAGCTCTGAGAAA 3939  
 1351 GlyArgProPheTyrSer 1356  
 3940 GGGAGGCCCTCTTATCTC 3957  
 seq\_name: gb\_AF467006  
 seq\_documentation\_block: 3933 bp mRNA linear PRI 07-FEB-2002  
 LOCUS AF467006 Homo sapiens partitioning-defective 3 protein splice variant f  
 DEFINITION (PARD3) mRNA, complete cds, alternatively spliced.  
 ACCESSION AF467006 GI:18568351  
 VERSION AF467006.1  
 KEYWORDS human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 3933)  
 Gao, L., Macara, I.G. and Joberty, G.  
 Multiple splice variants of PARD3 and of a novel related gene,  
 PARD3, produce functionally different proteins  
 Unpublished  
 JOURNAL 2 (bases 1 to 3933)  
 Gao, L., Macara, I.G. and Joberty, G.  
 Direct Submission  
 TITLE Submitted (10-JAN-2002) Center for Cell Signaling, University of  
 JOURNAL

FEATURES  
 source Virginia, PO Box 800577, Charlottesville, VA 22908, USA  
 Location/Qualifiers  
 1..3933  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..3933  
 /gene="PAR3"  
 /note="PAR3"  
 1..3933  
 /gene="PAR3"  
 /note="PAR3"  
 /note="PAR3; multi-PDZ domain containing protein;  
 alternatively spliced"  
 /codon\_start=1  
 /product="partitioning-defective 3 protein splice variant  
 f"  
 /protein\_id="AAU76046.1"  
 /db\_xref="id:18568352"  
 /translation="MKYTVCCRGRTRVVPCGGHMKVNSLIQAVTRRKALAKDPN  
 WIOVRLHEGDGILDDILCDVADDDKDLVAVDEDDPHHGSGTSSSTGTSP  
 IFGSELGNNNSAFQYQATSEIVTPSVLRANMPLHYRRSDPALIGLSTVSDSNF  
 SSEPSRKNPRTKMTTAGFLKQNTAGSPKTCRDKDENYRSILPDTSMNSNPFQDNA  
 RSSLSASHPMVGKLEKQEDDEGTDEEDNSRVEPYGHADTGLEHLPNSLDMVKLYE  
 VPNGCPILGIHVPPSARGRTIGLLYKRLKGGKAEHENLRENDCTVRIINDGLRN  
 RFEQAGHMFROAMRTPIITWYVPAANKEDYEQLSQSEKNNTYSRSPDSQYIDNR  
 SVNSGLHTVQAPRLNHPPEQIDSHSRLLPHSAHPSGKPPSAPASAPQNVFTTVSSG  
 YNTKIGRLNLIQKKTEGLGFSITSDVTIGGSAPYIKNLLPRGAILDGRLLKAG  
 DRLLENGVDVLGKQSEEVSLRSTKMEGVSLAVFROEDHPHRELKAEDEDIYLT  
 PDGREPLTFEVLNDGSAIGVSGKRSKENNADIGIFKSIINGSAAGKDLRL  
 VNDOLIAVNGSELGKTQNDAMETLRSMSPKRCNMQLIVARIISCNLKSQGS  
 PPGPELPTEIETLDDERKISHLSIGIEGLDSEPRNMLSLIMGKYQLSPIVNMED  
 DVIITIEDRLVPLPHLSQSSSSHDDVGYTAAGTAAKAIISDSACSSLPDVPD  
 VLAQREGGRQIADETKLTNTVDQKASPSHDVGSGLKSSLSLESLQTAVAEYTL  
 NGDIPIHRRPRIIRGRCCNESFRAAIDKSYDKPAVDDDEGMETLEEDTEERSNGR  
 ESVSASPDSOPSHLRQNNQNEKDKTDKDKDKTKKKKKRDKRDKKKKKKKGLK  
 GLGDMERGGKHKDKDKIKTKGKIKIOESTSEERIRMKQROERQAKRERERQAR  
 ERDYAEIODEFHTPECCDELMATGVSSTEGSALNARPOSPEGHMDALTYAQKPR  
 NSKSPVDSNRSTPSNHDRIQLRQEFQAKODEVEDERRITYSEQPMNAPRATQS  
 GRHSVSVOMORQROERESSQOARQYSSLPRQSRKRSKSSVSDSMQNSPGEGF  
 OSAKENPRYSYQSGRNGYLGCHFNARMLTEOIRLREORRKEQMKKPPSEGPS  
 NYDSYKRYODPEYAPPKGPFRRQDVPSPSQVARNLRLLQPERKGRPFYS"  
 814..1074  
 /gene="PAR3"  
 /note="Region: PDZ domain"  
 1378..1638  
 /gene="PAR3"  
 /note="Region: PDZ domain"  
 1726..1989  
 /gene="PAR3"  
 /note="Region: PDZ domain"  
 BASE COUNT 1164 a 899 c 1049 g 821 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 6754.00 Length: 1356  
 Ratio: 5.156 Gaps: 3  
 Percent Similarity: 96.608 Percent Identity: 96.608  
 alignment\_block:  
 US-09-757-781-2 x AF467006 ..  
 Align seg 1/1 to: AF467006 from: 1 to: 3933  
 1 MetLysValThrValCysPheGlyArgThrArgValValValProCysG1 17  
 |||||  
 1 ATGAAGTGAACCGCTGCTCGAGACGACCCGGGTGCTGCGCGTGGCG 50  
 |||||  
 17 yAspGlyHisMetLysValPheSerLeuIleGlnAlaValThrArgT 34  
 |||||  
 51 GAGAGCGCACATGAAAGTTTTCACCTCATCCACAGCGGTGACCGCT 100  
 |||||  
 34 yTrpGlySerAlaIleAlaLysAspProAsnTyrrTrpIleGlnValHisArg 50  
 |||||

101 ACCGGAAGCCATCGCCACAGATCCAAACTACTGTGATACAGTGCATCC 150  
 51 LeuGlnHisGlyAspGlyGlyIleLeuAspPheAspAspIleLeuCysAs 67  
 |||||  
 151 TTGGACATGTGAGATGAGAGAAATCTAGACTTGTATGATCATTTCTTGGCA 200  
 |||||  
 67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84  
 |||||  
 201 TGTAGACAGCATGAAGACACAGCTGTGTAGCTGTTGATGACGAGATC 250  
 |||||  
 84 roHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100  
 |||||  
 251 CACATTCACAGGAGTGATGCGACAGTGCAGTTCACACGGTACCCAGAGC 300  
 |||||  
 101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117  
 |||||  
 301 CCAGAGATATTGTGTAGTACCTTGCGCCAAATGCTCTACGCTTTTCA 350  
 |||||  
 117 nProTyrlAlaThrSerGluIleGluValThrProSerValLeuArgA 134  
 |||||  
 351 GCTTACCAACCAACAAAGTGAATGAGTGCACACCTTCAGTCTTCGAG 400  
 |||||  
 134 IaAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150  
 |||||  
 401 CAAATATGCCCTTCATGTTGACGCACTAGTGAACCCAGCTCTAATTGGC 450  
 |||||  
 151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167  
 |||||  
 451 CTCCTCCACTTGTGTCAGTATGATGATATTTCCTCTGAAAGACCTTCAAG 500  
 |||||  
 167 GlyAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184  
 |||||  
 501 GAAAAATCCACACAGCTGTCACAAACAGCTGGCTTCTCCTCAAGCAGACA 550  
 |||||  
 184 hTrAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrrArg 200  
 |||||  
 551 CTGTGGGAGTCTTAAACCTGCGACAGAAAGAAAGATGAAGAAACATACAGA 600  
 |||||  
 201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217  
 |||||  
 601 AGCCTCCCGCGGATCTACTAGTACTGGTCTTAACCAATTTCCAGACAGACA 650  
 |||||  
 217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234  
 |||||  
 651 TGTCTGCTGCTGCTGCTGAGTGCACAGTCCAAATGATGGGCAAGTGGCTGG 700  
 |||||  
 234 LuLysGlnGluGlnAspGluAspGlyThrGluGluGluAspAsnSerArgVal 250  
 |||||  
 701 AGAAACAAGAACAGATGAGATGGACAGAAAGATTAACAGTGTGT 750  
 |||||  
 251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267  
 |||||  
 751 GAACCTGTGGACATGCTGACACAGCGGTTTGGAGCATATACCACACTTTTC 800  
 |||||  
 267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284  
 |||||  
 801 TCTGGATGATATGATAAGCTGCTAGAGATCCCAACATGAGGAGGCGCTC 850  
 |||||  
 284 euGlyIleHisValValProPheSerAlaArgGlyGlyArgThrLeuGly 300  
 |||||  
 851 TGGGAATCCATGTAGTGGCTTTCAGTGTGCGAGGGGAGAAACCTGTGGGG 900  
 |||||  
 301 LeuLeuValLysArgLeuGluLysGlyLysIleLysAlaGlnHisGluAsnLe 317  
 |||||  
 901 TTATTATTAGTAAACAGATTGAGAAAGTGTGAAGCTGAACATGAAGAAACT 950  
 |||||  
 317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArgA 334  
 |||||  
 951 TTTTCGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1000  
 |||||  
 334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350  
 |||||  
 1001 ATGAGAGATTGTAACAGCAACATATATGTTTGGCAAGCCATGCGGTACA 1050  
 |||||

351 ProIleIleTrpPheHisValProAlaAlaAsnLysGluIleTrpG 367  
 |||||  
 1051 CCCATCATTTGGTTCCATGTGGTCTTCGACGAATAAAGAGCATATTA 1100  
 367 uGlnLeuSerGlnSerGluLysAsnAsnTrpTrpSerSerArgPheSerP 384  
 |||||  
 1101 ACACTATTCCTCCAAAGTAGAGAACAAATTACTATTCACGCCGTTTAAGCC 1150  
 384 roAspSerGlnTrpIleAspAsnArgSerValAsnSerAlaGluLeuHis 400  
 |||||  
 1151 CTGACAGCCAGATATTTGACACAGAGAGTGTGACAGTCTCAGGGCTTCAC 1200  
 401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417  
 |||||  
 1201 ACGGTGCAAGAGCACCCCGACTGAACACCGCGCTGAGCAGATGACTCG 1250  
 417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434  
 |||||  
 1251 TCACCTCAAGACTACTCATAGCGCACACCCCTCGGGAAACACCATCCG 1300  
 434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450  
 |||||  
 1301 CTCACGCCCTCGGCACCTCAGAAATGTAATTAGACACTGTAAAGCATGCT 1350  
 451 TrpAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467  
 |||||  
 1351 TATACACCAAAAATAAGGAGAGGCTTAAATCCAGCTTAAGAAGAG 1400  
 467 TrpGlnGluLysGlyPheSerIleThrSerArgAspValThrIleGlyG 484  
 |||||  
 1401 TACACAAGGTTGGGATTCAGCATCACCTCCACAGATGTAAACAATAGGTG 1450  
 484 LySerAlaProIleTrpValLysAsnIleLeuProArgLysAlaAlaIle 500  
 |||||  
 1451 GCTACGCTCCCATCTATGTGAAAAAATCTCCCGCGGGGCGGCATTT 1500  
 501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGlnValAsnGlyVal 517  
 |||||  
 1501 CAGATGGCCGACTTAAGCGACGAGACAGACTTATAGAGGTAAATGAGAT 1550  
 517 laSpleuValGlyLysSerGlnGluIleValSerLeuLeuArgSerT 534  
 |||||  
 1551 AGATTGTGTGGGCAAAATCCCAAGAGCAAGTTGTTCCGCTTGAGAAAGCA 1600  
 534 hrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550  
 |||||  
 1601 CCAAGATGAGAGCACTGTAGCCCTTCTGTGCTTCGCCAGAAAGACGCC 1650  
 551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567  
 |||||  
 1651 TTCACCCCAAGGGAAGCTG..... 1668  
 567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrA 584  
 |||||  
 1669 .....AAAGCAAGAGATGAGGATTTGTTCTTACACCTGATGGACACA 1711  
 584 rGluIlePheLeuThrPheGluValProLeuAsnAspSerGlySerIleAgly 600  
 |||||  
 1712 GGGAAATTTCTGACATTTGAAGTCCCACTTAATGATTCAGAGATCTCAGAGC 1761  
 601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617  
 |||||  
 1762 CTTGGGTGTAGTGTCAAAAGTAAACCGGTCAAAAGAAACCAAGCAGATT 1811  
 617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634  
 |||||  
 1812 GGGAAATCTTTGCAAGTCCATTTATTAATGAGAGACACCATCTAAAGATG 1861  
 634 LyArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerLeu 650  
 |||||  
 1862 GAAGCTTCGGGTGAATGATCATCACTGATAGCAGTAAATGGAATCCCTG 1911

651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667  
 |||||  
 1912 TTGGCCAAAGACAAACCAAGATGCCATGGAACCTTAAGAAAGTGTATGTC 1961  
 667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684  
 |||||  
 1962 TACTGAAGGCAATTAACGAGAAATGATCCAGCTTATTTGTTCAAGAGCA 2011  
 684 leSerLysCysAsnGlnLeuLysSerProGlySerProProGlyProGlu 700  
 |||||  
 2012 TAACCAAGTGCATGAGACTGATAGCTACCTGGAGACCCCTGGACCTGAG 2061  
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717  
 |||||  
 2062 CTGGCAATTAACAGCGCTTGATGATAGAGAACGAAGATTTCCCATTC 2111  
 717 rLeuTrpSerGlyIleGlnGlyLeuAspGluSerProSerArgAsnAla 734  
 |||||  
 2112 CCCTACAGTGGGATTTGAGGGGCTTGATGATCCCCACAGCAAAATGCTG 2161  
 734 laLeuSerArgIleMetGlyLysSerGlyLysTrpGlnLeuSerProThr 750  
 |||||  
 2162 CCCTCAGTAGATATATG.....GTTAAATACAGCTGTCCCTTACA 2202  
 751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeu 767  
 |||||  
 2203 GTCAATATGCCCCCAATGATGACACTGTCAATTATAGAAATGACAGGTTGCC 2252  
 767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784  
 |||||  
 2253 AGTGGTCTCCACATCTCTGACACAGTCTTCCACGCTCCCAAGATG 2302  
 784 spValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIle 800  
 |||||  
 2303 ATGTGGGTTTGTGTGACGGCAGATGCTGTACTTGGGCCAAGGCTCAATC 2352  
 801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817  
 |||||  
 2353 AGTGATTCAGCCGACTGCTCTTTGAGTCCAGATGTTGATCCAGGTTCTTGC 2402  
 817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834  
 |||||  
 2403 TTTTCAACAGAGAGATTTGGACGTCAG..... 2430  
 834 ySGLnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850  
 |||||  
 2430 ..... 2430  
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867  
 |||||  
 2431 .....ATAGCTGAGAGACTTAACCTCAATACAGTGA 2462  
 867 PASpGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884  
 |||||  
 2463 TGACCAGAAACAGAGTTCTCCACACAGATGTTGGGCTTCCCTGGGTC 2512  
 884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900  
 |||||  
 2513 TGAAGAATCAAGCTCTTGGAGAGTGTGCAGACCGGAGTGTGGCAGGGT 2562  
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleArg 917  
 |||||  
 2563 ACTTGAATGGGATATTTCTTCCATGTCGTCACGGCCGGGATATATCAG 2612  
 917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTrpA 934  
 |||||  
 2613 AGGCAGGGCGATGCAATGAGAGCTTCAGAGCTGCCATCGCAAAATCTTATG 2662  
 934 sPLysProAlaValAspAspAspAspGlyLysIleGluThrLeuGluGlu 950  
 |||||  
 2663 ATAAACCCCGGTAGATGATGATGAAGCATGGAAGCACTTGGAGAGAA 2712  
 951 AspThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlase 967

```

|||||
2713 GACACGAGAAAGTTCAGATCAGGAGAGAGCTGTATCCACAGCCAG 2762
967 TASPGLNProSerHisSerLeuGluArgIleMetAsnGlyAsnGlnGlu 984
2763 TGATACGCTTCCCTCTCTGTGAGAGACAAATGATGAAACCAAGAGA 2812
984 ysgIyAspLystrHisParGlyLysAspLysThrGlyLysGlyLys 1000
2813 AAGGTGATAGACTGATAGAAAAAGAGATAAACTGGAAAAAGAAAGAG 2862
1001 LysAspArgAspLysGlyLysAspLysMetLysAlaLysGlyMetLe 1017
2863 AAAGATAGAGATAGAGGAAGAGATAAATGAAAGCCAAAGAAAGGAAGCT 2912
1017 uLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAsp 1034
2913 GAAGGCTTGGGAGACATGTTCAAGTGTGGCAAAACATCGAAAAGATGACA 2962
1034 ysIleGlyLysThrGlyLysIleLysIleGlnGlnSerPheThrSerGlu 1050
2963 AGATTGAGAAACGGGTAAATATAATACAGGAATCCTTTACATCAGAA 3012
1051 GlnGluArgIleArgMetLysGlnGlnGlnGluArgIleGlnAlaLysTh 1067
3013 GAGAGAGAGTACGAATGAAGAGAGAGAGAGAGATTCAGCCAAAC 3062
1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyrAlaGlnIleG 1084
3063 TCAGAGATTTAGGGAGACAGACAGACAGCTCGAGACGCTAGCTGAAATTC 3112
1084 InAspPheHisArgThrPheGlyCysAspAspGluLeuMetTyrGly 1100
3113 AAGATTTCATCGAGACATTTGGCTGTGATGATGATTAAATGATAGGGGA 3162
1101 ValSerSerTyrGlnGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117
3163 GTTCTCTTATGAAAGTTCATGAGCTCCTACAGCTAGACCTCAGACCCC 3212
1117 cArgGlnGlyHisMetMetAspAlaLeuTyrAlaGlnValLysLysPro 1134
3213 ACAGAAAGGCAATATGATGATGCTTTGTAATGCCAAGTCAAGAAAGCCGC 3262
1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
3263 GGAATTCCAAAACCTCACCTGTAGACAGTAAACAGATCACTCCATGCAAT 3312
1151 HisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAs 1167
3313 CATGATCGATACAGCGCTGAGGCAAGAAATTCAGCAAGCAAAAGCAGAGA 3362
1167 pGluAspValGluAspArgArgArgThrTyrSerPheGlnGlnProTyrP 1184
3363 TGAAGATGTAGAGATGCTGCGCGGACCTATAGTTTGAGCAACCCCGGC 3412
1184 roAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200
3413 CGAAGCGACGCGCGGCGAGCGAGCGGCGACACTCGTGTCGCTGAGAG 3462
1201 ValGlnMetGlnArgGlnArgGlnGlnGluArgLysSerGlnGlnAl 1217
3463 GTGCGATGTCAGCGCGGCGAGCGGAGAGAGCGCGAGAGCTCCGACAGGC 3512
1217 aGlnArgGlnTyrSerSerLeuProArgGlnSerArgLysAsnAlaSer 1234
3513 CGAGCGGCAATACAGCTCTCTGCTCGCAAAAGAGAAAAATGCCAGCT 3562
1234 eValSerGlnAspSerTyrpGlnGlnAsnTyrSerProGlyGlnGlyPhe 1250
3563 CGGCTCTCCAGAGACTCTTGGGAGCAGAACTACTCCCTGGGGAAGGCTTC 3612
1251 GlnSerLysGlnAspProArgTyrSerSerTyrGlnGlySerArgAs 1267
|||||

```

```

3613 CAGAGTCCCAAAAGAAACCCAGGTACTCCAGCTACCAAGGCTCCAGGAA 3662
1267 nGlyTyrLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGlnThrG 1284
3663 CGGCTACTTGGGAGACATGCTTCAACGCCAGGGTATGCTGAAACTC 3712
1284 lngIleuLeuArgGlnGlnGlnArgArgLysGlnGlnMetLysLys 1300
3713 AGGAGCTCTTCGCCAGAGAACAGAGCGGAGGAGAGAGAGATGAAGAG 3762
1301 GlnProProSerGlnGlyProSerAsnTyrAspSerTyrLysValGln 1317
3763 CAGCTCTCTCCGAGGGGCCAGCACTATGACTCGTATAGAAAGTCCA 3812
1317 nasPProSerTyrAlaProProlysGlyProPheArgGlnAspValProp 1334
3813 GGACCCCAAGTACGCCCTCCCAAGGGGCCCTTCCGGCAAGATGTGCC 3862
1334 roSerProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGluLys 1350
3863 CTTCCCTTCACAGGTTGCGAGGCTGAACAGACTTCAGACTCTGAGAAA 3912
1351 GlyArgProPheTyrSer 1356
3913 GGGAGGCCCTTCTATTC 3930

seq_name: gb_pr:AF332593

seq_documentation_block: 3822 bp mRNA linear PRI 16-NOV-2001
LOCUS AF332593
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein long
variant b mRNA, complete cds, alternatively spliced.
ACCESSION AF332593
VERSION AF332593.1 GI:14579310
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3822)
Fang,C.M. and Xu,Y.H.
Down-regulated expression of atypical PKC-binding domain deleted
asip isoforms in human hepatocellular carcinomas
Cell Res. 11 (3), 223-229 (2001)
JOURNAL MEDLINE 21520080
PUBMED 11642408
REFERENCE
2 (bases 1 to 3822)
Fang,C. and Xu,Y.
Exon/Intron Structure And Splicing Variants of A Novel Human
Polarity Gene, hASIP
Unpublished
JOURNAL 3 (bases 1 to 3822)
AUTHORS Fang,C. and Xu,Y.
TITLE Direct Submission
JOURNAL Submitted (29-DEC-2000) Laboratory of Molecular and Cellular
Oncology, Institute of Biochemistry and Cell Biology, Shanghai
Institutes for Biological Sciences, Chinese Academy of Sciences,
320 Yue-Yang Road, Shanghai 200031, P. R. China
Location/Qualifiers
FEATURES
source
1..3822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
1..3822
/Note="ASIPb; alternatively spliced; similar to GenBank
Accession Number AF196185"
/codon_start=1
/product="atypical PKC isotype-specific interacting
protein long variant b"
/protein_id="AAK69193.1"
/db_xref="GI:14579311"
/translation="MKYIVCFGRTRVYVFCGSGIMKRVFSLIQAVTRKALAKDPNY

```

WIOVHLEHGDCGIIIDLDIICDVADDKRLVAVFEDQPHHGDTASSTGTQSPRE  
IFGSELGTNNVSAFQTOATSEIYTPSVLRANMPLHVRSSDALIGLSTVSDSNF  
SSEPSRKNPTNMTSTTAGLKNONTAGSPKCTDRKKNEDENYRSLPDTNMSVOPORDVA  
RSLSLSHPMVGMLEKQEQDQDGTEDNSRREPKGADTLEHLPNLSLDMKYLVE  
VPDGPPLGIHVVPFARGGRILGLIVLRLEGGAEBENI.PRENDICVLRINDDLN  
RPEOQHMFROAMRTPIIMFHVPAANKKEQYEQJ.SQSKNNYSSRPSPOXYIDNR  
SYNSAGLHTVORAPRLNHPPEOIDSLSRLPHSAHSGKPSAPASAPONVSTYSSG  
YNTKIGKRLNIOIKKGTGLGFSITSRDVYIGGSAPIYKVKLIPKRAIODGRLKGG  
DLRLVKNGVLVKRSQSEVYSILRSTMEGTYSKLVFROEDAFHRELKADIEDVLT  
PDGTREFLUTEVPLNDSGSLGCVSYKNSKENHADLGIYFKSIIINGGASKDBRLK  
VNDOLIVNGESILGKTNDQAMETLRPSTBENKNGMIQLIIVARIISKCNELKSPGS  
PPGDELLETALDDREKRISHSLTSGIBELDSPRNALSTRMGKYOLOSPVNNPQD  
DVTIIEDDRPLVPLPHLSDQSSSSSHDDVGFYTPADGAKAKAISDASCSLSPVDV  
VIAFORBEGFROIADETKLTNTVDQKAGSPSRVGVSLGKSSLSLQTOYVAEYTL  
NGDIPFHNRPRIIRGRGCNBSFPAIDKSYDPKAVDDDEGMEILEPTTESSRSGR  
ESVSTASPOSSHSLEROMNGNOEKDGTDRKKDKTKERKKDKDKDKDKKMK  
GLDMEFRLQAKTRERROARENDYAEIODFHTFGCDDIEMATGVSYSKMAIAR  
POSREGHMDALTAQYKKNPSNKPSPVDSNRPSPNHDIORLROEFQOAKODEVE  
DRRTYSEQWPMPNAPATQSGRHVSVEVOMOROEERESSQQAORQYSLPQSR  
KNASVSODSEONTSPEGFOSKAKENRYSYSGSGYLGHGHNARVMELOLL  
ROEORRKEQMKOPPESEGPSNYSYKRVQDPSIAPPKGFRQDVPSPSOVARLNLRL  
QPREKRPYS\*

BASE COUNT 1114 a 885 c 1019 g 804 t  
ORIGIN

alignment\_scores: Length: 1356  
Quality: 6537.50 Gaps: 4  
Ratio: 5.136  
Percent Similarity: 93.879 Percent Identity: 93.879

alignment\_block:  
US-09-757-781-2 x AF332593 ..

Align seg 1/1 to: AF332593 from: 1 to: 3822

1 MetLeuValThrValCysPheGlyArgThrArgValValProCysG1 17  
|||||  
1 ATGAAAGTACGCGTGTCTCGACGACCGCGGTGCTGCGCGG 50  
17 yAspGlyHisMetLeuValPheSerLeuIleGlnGlnAlaValThrArgT 34  
|||||  
51 GGACGGCCACATGAAAGTTTTCAGCTCATCCAGCGGCGTGACCGGCT 100  
34 yTArgGlyAlaIleAlaValAspProAsnTyrTrpIleGlnValHisArg 50  
|||||  
101 ACCGGAAGGCCATCGCCAAAGATCCAAACTGAGTACAGGTGCAATCCG 150  
51 LeuGlnHisGlyAspGlyIleLeuAspLeuAspAspIleLeuCysAs 67  
|||||  
151 TTGGAACTGGAGATGGAGAACTAGACCTTGATGACATTCCTTGTGA 200  
67 pValAlaAspAspIlyAspArgLeuValAlaValPheAspGluGlnAsp 84  
|||||  
201 TGTAGACAGACATAAAGACAGCTGTAGCAGTGTGGATGAGAGAGATC 250  
84 roHisHisGlyIlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100  
|||||  
251 CACATCACGAGGTGATGGCACACAGTCCACGCGGTACCCAGAGC 300  
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117  
|||||  
301 CCAGAGATATTTGTGTAGCTGTGGCCACCAACAATGTCTGACGCTTCA 350  
117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValIleuArg 134  
|||||  
351 GCGTTACCAAGCAACAAGTGAATGTGAGGTTCACACCTTCAGTCCCTCG 400  
134 laAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150  
|||||  
401 CAAATATGCGCTTCATGTTGAGCGCAGTAGTACCCAGCTCTGAATTTGC 450  
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167

|||||  
451 CTCGCCACTCTCTGCAGTGTAGTAAATTTTCTCTGAAAGACCTTCAAG 500  
167 GlyAspProThrArgTrpSerThrThrAlaGlyPheLeuGlyGlnAsnT 184  
|||||  
501 GAAAAATCCCAACGCGTGGTCAACAACAGGTGCTCTCCAAAGCAGMAA 550  
184 hrAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200  
551 CTGCTGGGAGTCCTAACCTCGCACAGAAAGATGAAATCTACAA 600  
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217  
601 ACCCTCCCGCGGAGTACTAGTAACTGTGTAACCAATTTGAGAGACAA 650  
217 naLaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234  
|||||  
651 TGCTCGCTGCTCTGAGTGGCCAGTCCACCCAAATGATGGGCAATGGCTG 700  
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250  
701 AGAAACAAGACAGATGAGATGGGACAGAGAGGATTAACAGTCTGCTT 750  
251 GluProValGlyHisAlaThrGlyLeuGlnHisIleProAsnPheSe 267  
751 GAACCTGTGGACATGCTGACACAGGTTTGGAGCATATACCACTTTTC 800  
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyGlyProL 284  
801 TCTGATGATATGATGAAGCTCGTAAGAGTCCCAACGATGGAGGGCTTC 850  
284 euGlyIleHisValValProPheSerAlaArgGlyLysArgThrLeuGly 300  
851 TGGGATCCATGTAGTGCCTTTCAGTGTGTCGAGCGCGCAACCTCGGG 900  
301 LeuLeuValLysArgLeuGluLysGlyGlyLysAlaGlnHisGlnAsn 317  
901 TTATTAGTAAACGATTTGGAGAAAGGTGGTAACCTGAACATGAAATCT 950  
317 urPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334  
|||||  
951 TTTTCTGAGAAATGATTTGATTCAGGATTAATGATGGGACCTTGAA 1000  
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350  
1001 ATGAGACATTTGAAACAGCAACATATGTTTCGCAAGCCATCGTACA 1050  
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnTyrG1 367  
1051 CCCATCATTTGGTTCATGTGGTCCGACGCAATTAAGACAGCATATGA 1100  
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSer 384  
1101 ACAACTATCCCAAGGTGAGAAACAAATTAATTAATTAACCGCTTTAGCC 1150  
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerIleGlyLeuHis 400  
1151 CTGACAGCGCATATATGACAAACAGAGTGTGAACAGTGGAGGCTTCA 1200  
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAsp 417  
1201 ACGGTGCGAGAGACACCCGACTGAACACCCCGCTGAGAGATAGATTC 1250  
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSer 434  
1251 TCACCTCAAGACTACTCTATACGCAACACCCCTCGGAAACCAACATCC 1300  
434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450  
1301 CTCACACCTCGGCACCTCGAATGATTTTATGACAGCTGTAACGATGCT 1350  
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG1 467



```

1351 TATTAACCAAAAAATAGGCAAGAGCTTAAATATCCAGCTTAAGAAAGC 1400
467 yThrgluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
|||||
1401 TACAGAAAGGTTGGATTGCATCTCAGACTTCCAGAGATGTAAACAATAGGTG 1450
484 ySerIleProIleTyValValAsnIleLeuProArgGlyAlaIle 500
|||||
1451 GCTCAGCTCCAAATCTATGTGAAAAACATTCCTCCCGGGGGGGCGCATTT 1500
501 GlnAspGlyArgLeuValAspArgLeuIleGluValAsnGlyVal 517
|||||
1501 CAGGATGGCCGCACTTAAGCAGCAGACAGACTTATAGAGSTAAATGGAGT 1550
517 LaspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSer 534
|||||
1551 AGATTATAGTGGCAAAATCCCAAGAGAGAGTTGTTGCTGTGAGAAAGCA 1600
534 hrlYsMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
|||||
1601 CCAAGATGGAGAGAGACTGTGAGCCTTCTGCTTTCGCCAGGAAGAGCGCC 1650
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProly 567
|||||
1651 TTCACCCCAAGGGAGCTG..... 1668
567 sgluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584
|||||
1669 .....AAAGCAGAAAGATGAGATATTGTTTCTTACACCTGATGGCA 1711
584 rglGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
|||||
1712 GGGAAATTTTCGACATTGTGAAGTCCCACTTATATGATTCAGATCTGCAGGC 1761
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
|||||
1762 CTTCGGTGTCAAGTCAAGGTAAACCGGTCAAAAGAGACACCGACAGATTT 1811
617 uGlyIlePheValLysSerIleLeaSnGlyAlaAlaSerLysAspG 634
|||||
1812 GGGAAATCTTGTCAAGTCCATATTATTAATGAGAGCAGCATCTAAAGATG 1861
634 LyrArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650
|||||
1862 GAAGGCTTCGGGTGAATGATCAACTGATACAGTAATGAGAAATCCCTG 1911
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
|||||
1912 TTGGCCAGACCAACCAAGATGCCATGGAAACCTTAAGAAAGTCTATGTC 1961
667 rThrgluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684
|||||
1962 TACTGAAGGCAATTAACGAGGAATGATCCAGCTTATGTTGCCAAGAGAA 2011
684 IeSerLysCysAsnGluLeuLysSerProGlySerProGlyProGlu 700
|||||
2012 TAAGCAAGTGCATAGCTGAAGTCACTGAGCCCTCGGACCTGAG 2061
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
|||||
2062 CTGCCATTAAACACAGCGTTGGATGATAGAGAACAGAAATTTCCCATTC 2111
717 rLeuTySerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
|||||
2112 CCTCTACAGTGGGATTGAGGGGCTGTGATGAATGCCCAAGCAAAATGCTG 2161
734 lalauserArgIleMetGlyGluSerGlyLysTyrglnLeuSerProThr 750
|||||
2162 CCCTCACTAGATAAATG.....GGTAAATACCAAGCTGCCCTACA 2202
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767
|||||
2203 GTGAATATGCCCAAGATGACACTGTATTAATAGAGATGACAGGTTGCC 2252
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerSerHisAspA 784
|||||
2253 AGTGCCTCTCCACATCTCTCTGACACAGTCTTCCAGCTCCCATGATG 2302
784 spValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800
|||||
2303 ATGTGGGGTTTGTACGCGCAGATGCTGTACTTGGGCGCAAGCGTGCATTC 2352
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
|||||
2353 AGTCATTACAGCCGCTCTCTCTTTCAGTCCAGATTTGATTCAGATTCTTGC 2402
817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrl 834
|||||
2403 TTTTCAACGAGAAAGATTTGGAGCTCAG..... 2430
834 ysgInPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
2430 ..... 2430
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
2431 .....ATAGCTGACGAGACTAAACTCAATACAGTGA 2462
867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
2463 TGACACCAAAAGCAGATTCTCCACAGACAGATGGTGGTCTCCCTGGGTC 2512
884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900
2513 TGAAGAGTCAAGCTCATTTGGAGAGTCTGCAGACCGCAGTTCGAGAGTG 2562
901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleAr 917
2563 ACTTTGAATGGGATTTCTCTTCCATCTGTCACGGCCGGGATTAATCAG 2612
917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyra 934
2613 AGCGAGGGGATGCAATGAGAGACTTCAGAGCTGCATCGACAAATCTTATG 2662
934 spLysProAlaValAspAspAspAspGluGlyMetGluThrLeuGluGlu 950
2663 ATTAACCCGCGGTAGATGATGATGATGAAGCATGGAACCTTGGAAGA 2712
951 AspThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlaSe 967
2713 GACACAGAGAAATTTCAAGATCAGGAGAGAGTCTGTATCCACAGCCAG 2762
967 rAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlyAsnGlnGlu 984
2763 TGATCAGGCTTCCCACTCTCTGGAGAGACAAATGAATGGAACCAAGGA 2812
984 ysgIAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000
2813 AAGGTGATTAACACATGATGAGAAAAAGATTAACCTGGAAGAAAAAGAA 2862
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
2863 AAAGATAGAGTTAAGAGAGATTAATAATGAAGCCAAAGAAAGGAAATCTT 2912
1017 uLysGlyLeuGlyLysMetPheArgPheGlyLysHisArgLysAspAspL 1034
2913 GAAGGCTTGGAGACATGTTCT..... 2934
1034 yslIleGluLysThrGlyLysIleLysIleGlnGluSerPheThrSerGlu 1050
2934 ..... 2934
1051 GluGluArgIleArgMetLysGlnGluGlnArgGlyIleGlnAlaLysTh 1067
2935 .....AGGATTCAAAGCCAAAC 2951

```



misc.feature /note="Region: PDZ domain 2"  
1594..1851  
/gene="PAR3"  
/note="Region: PDZ domain 3"  
BASE COUNT 1121 a 867 c 1016 g 797 t  
ORIGIN

alignment\_scores:  
Quality: 6485.00 Length: 1356  
Ratio: 5.122 Gaps: 4  
Percent Similarity: 93.363 Percent Identity: 93.363

alignment\_block:  
US-09-757-781-2 x AF252293 ..

Align seg 1/1 to: AF252293 from: 1 to: 3801

```
1 MetLysValThrValCysPheGlyArgGlyArgValValProCysG1 17
|||||
1 ATGAAAGTGCACCGTGCCTTCGAGACGACCGGGTGGTCCGCCGTGCG 50
17 YASPGLYHISMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34
|||||
51 GGACGGCCACATGAAAGTTTTCAGCCCTCATCCAGACGGCGGTACCCGCT 100
34 YRArgLysAlaIleAlaLysASPProAsnTyrTrpIleGlnValHisArg 50
|||||
101 ACCGGAAGGCCATCGCCAAAGATCCAAACTACGTGATACAGGTCCATCGC 150
51 LeuGluHisGlyAspGlyGlyIleLeuAspLeuAspAspIleLeuCysAs 67
|||||
151 TTGGACATGAGATGAGCAATACATCTGACCTTGATGACATCTCTTGTA 200
67 PValAlaAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84
|||||
201 TGTAGCAGAGATTAAGACAGACTGGTAGCAGTGTGGATGATGAGGATC 250
84 rHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
|||||
251 CACATACCGAGGTGATGGCAGCACAGTCCAGTCCACGGGTACCCACAGC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
|||||
301 CCAGAGATATTGTGTAGTGTGGCACCACAAATGTCTCAGCCCTTCA 350
117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
|||||
351 GCCTTACCAAGACAAGTGAATTTGAGTCCACCTTCAGTCCCTTGAG 400
134 LaAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
|||||
401 CAATATGCTCTTCATGTTCCAGACGAGTAGTGACCCAGCTCTAATTGGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
451 CTCTCCACTCTCTGAGTGAATTAATTTTCCCTGAAGAGCCCTTGAC 500
167 GlyAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184
501 GAAATAATCCCAAGCGTGTCAACACAGCTGGCTCTCAAGACGAGACA 550
184 hAlaArgSerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
551 CTGCTGGAGATCCGTAACCTGGACAGGAAG..... 582
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217
582 ..... 582
217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeuG 234
582 ..... 582
```

```
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
|||||
583 .....GATGAGATGGGACACAGAAAGCATTAACAGTCGTCTT 618
251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPheSe 267
|||||
619 GAACCTGTGGACATGCTGACACGGGTTTGGAGCATTAACCACTTTTC 668
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyGlyProL 284
669 TCTGATGATATGTTAAAGCTCGTAGAAGTCCCAAGAGTAGAGGCGCTC 718
284 euGlyIleHisValAlaProPheSerAlaArgGlyGlyArgThrLeuGly 300
719 TGGGAATCCATGTAGTCCCTTTCAGTCTCTGAGGCGGACAGACCTTGCG 768
301 LeuLeuValLysArgLeuGluLysGlyGlyLysAlaGluHisGluAsnLe 317
769 TTATTACTAAACGNTTGGAGAAAGGTGTAAAGCTGAACATGAAATCT 818
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
819 TTTTCGTAGATGATTTGCAATGTCAAGATTAATGATGGGACCTTGAA 868
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
869 ATGAAAGATTTGAACACAGCAACATATGTTGCCCAAGCATCGCTACA 918
351 ProIleIleTrpPheHisValAlaProAlaAlaAsnLysGluGlnTyrG1 367
919 CCCATCATTTGGTTCATGTGTCTCCAGCAAAATGAAGCGCATGTGA 968
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384
969 ACAACTATCCCAAGATGAGAAACAATTACTATCAAGCGTTTATGCC 1018
384 roAspSerGlnTyrIleAspAsnArgSerValaAsnSerAlaGlyLeuHis 400
1019 CTGACACCGCATATATTGACAAACAGAGTGTGAACAGTCCGCTTCAC 1068
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
1069 ACGGTGCAGAGAGCACCCCGACTCAACACCCCGCTGAGCAGATAGACTC 1118
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSera 434
1119 TCACCTCAAGACTACCTCATAGCGCACACCCCTCGGAAACACCATCCG 1168
434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1169 CTCACGCTCGGCACCTCAGAAATGATATTAGTACAGCTGAAGCAGTGT 1218
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG1 467
1219 TATTAACACCAAAAAAATAGGCMAAGGCTTAATATCCAGCTTAAGAAAG 1268
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484
1269 TACGAAGGTTTGGATTCAACATCACTCCAGAGATGTAACAATAGGTG 1318
484 lYserAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
1319 GCTCAGCTCCCAATCATGTGAAAAAANTTCCTCCCGGGGGGGCCCATTT 1368
501 GluAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValaAsnGlyVa 517
1369 CAGATGGCCGACTTAAGCGAGAGACAGACTTATAGAGTAAATGGAGT 1418
517 lAspLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSerT 534
1419 AGATTTAGTGGCAAAATCCCAAGAGGAAGTTTTCGCTGTTGAGAGACA 1468
```

534 hrlysmetClnGlyThrValSerLeuValPheArgGlnIuAspAla 550  
 |||  
 1469 CCAAGATGGAGAACTGTGAGCTTCTGCTTCCGCAAGAAACCC 1518  
 551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567  
 |||  
 1519 TTCACACCAAGGAAGTCTG..... 1536  
 567 sglutThrValAlaGluAspGluAspIleValLeuThrProAspGlyThra 584  
 |||  
 1537 .....AAAGCAGAGATGAGATATGTCTTCTACACCTGATGGCACA 1579  
 584 rgluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600  
 |||  
 1580 GGGATTTTCGACATTTGAAGTCCCACTTAATGATTCAGGATCTGCAGCG 1629  
 601 LeuGlyValSerVallysgIysAsnArgSerLyGluAsnHisAlaAspLe 617  
 |||  
 1630 CTTCGTCTCAGTGTCAAAAGGTAAACCGGTCAAAAGACACCCAGCATTT 1679  
 617 uGlyIlePheVallySerIleIleAsnGlyAlaAlaSerLyAspG 634  
 |||  
 1680 GGGATTTTCGACATTTGATCCATTTATAGAGAGCAGCATTAAGATG 1729  
 634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyIuSerLeu 650  
 |||  
 1730 GAAAGCTTCGGGTGTAATGATCACTGATAGCAATTAATGGAGAAATCCCTG 1779  
 651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667  
 |||  
 1780 TTGGGCAGACAAACCAAGATGCCATGGAAACCTTAGAAGGTCTATGTC 1829  
 667 rThrGluGlyAsnLyArgGlyMetIleGlnLeuIleValAlaArgArg 684  
 |||  
 1830 TACGGAAGCAATTAACGAGAGATGATCCAGCTTATTTGTCAGAGGAAA 1879  
 684 IeSerLyCyAsnGlnLeuLysSerProGlySerProGlyProGly 700  
 |||  
 1880 TAAAGAACTGCATGAGCTGAACTGACCTGGAGACCCCTGGAGCCGAG 1929  
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717  
 |||  
 1930 CTGCCCATTTGAACAGCGTTGATGATAGAGACGAAATTTCCCATTC 1979  
 717 rLeuTyArgGlyIleGlyGlyLeuAspGlySerProSerArgAsnAla 734  
 |||  
 1980 CCTGTACAGTGGGATGAGGGCTTGATGAATCCCAAGCAGAAATGCTG 2029  
 734 lAleuserArgIleMetGlyIuSerGlyLysTyArgInLeuSerProThr 750  
 |||  
 2030 CCTCAGTAGATTAATG.....GGTAAATACCACTGTCCTCCCTACA 2070  
 751 ValAsnMetProGlnAspAspThrValIleIleGlnAspAspArgLeuPr 767  
 |||  
 2071 GTGAATATGCCCCCAAGATGACACTGCTATTATAGAAATGACAGGTTC 2120  
 767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784  
 |||  
 2121 AGTGTCTCTCCACATCTCTGACACAGTCTCTCCAGCTCCCATGATG 2170  
 784 spValGlyPheValThrAlaAspAlaGlyThrTrpAlaIysAlaIle 800  
 |||  
 2171 ATGTGGGGTTTGTGAGCGCAGATGCTGTACTGTGGGCCAAAGCTGCATC 2220  
 801 SerAspSerAlaAspCySerSerLeuSerProAspValAspProValLeuAl 817  
 |||  
 2221 AGTGATTCAGCCGACTGCTCTTTGAGTCCAGATGTGATCCAGTTCTTGC 2270  
 817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGlyLysArgThrL 834  
 |||  
 2271 TTTTCAACGAGAAAGATTTGACGTGAG..... 2298  
 834 ysgInPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850

2298 ..... 2298  
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867  
 |||  
 2299 .....ATAGCTGACGAGACTTAACTCAATCACTGAG 2330  
 867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884  
 |||  
 2331 TGACCAGAAAGCAGGTCTCCACAGAGATGTGGGTCTCTCCCTGGGTC 2380  
 884 euLySlySerSerSerLeuGlyuSerLeuGlnThrAlaValAlaGluVal 900  
 |||  
 2381 TGAAGAAGTCAAGCTCGTTGGAGAGTCTCCAGACCCGAGTTGCCGAGGTG 2430  
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleArg 917  
 |||  
 2431 ACTTGAATGGGATATTCCTTCTCATCTGTCACAGCGCCGCGAATACAG 2480  
 917 gGlyArgGlyCyAsnGlyuSerPheArgAlaAlaIleAspLySerTyra 934  
 |||  
 2481 AGCGAGGGGATGCAATGAGACTTCAGAGCTGCCATCGACAAATCTTATG 2530  
 934 sPLysProAlaValAspAspAspGlyuGlyMetGluThrLeuGluGlu 950  
 |||  
 2531 ATMAACCCGCGGTAGATGATGATGATGATGAGGCAATGATGATGATGATG 2580  
 951 AspThrGluGlyuSerSerArgSerGlyArgGlyuSerValSerThrAlaSe 967  
 |||  
 2581 GACACAGAAAGAAAGTTCAAGATCAAGGAGAGAGCTGTATCCACAGCCAG 2630  
 967 rAspGlnProSerHisSerLeuGluArgGlnMetLysnGlnGlnGlu 984  
 |||  
 2631 TGATCAGCTCTCCACTCTCTGTGAGAGACAAATGATGATGAAACCAAGGA 2680  
 984 yGlyLysPlyThrAspArgLysLysAspLyThrGlyLysGlyLys 1000  
 |||  
 2681 AAGGTGATAGACTGATGACAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2730  
 1001 LysAspArgAspLyGlyLysAspLyMetLysAlaLysLysGlyMetLe 1017  
 |||  
 2731 AAAGATAGAGATTAAGGAGAAAGATTAAGAAAGCAAGAAAGGAAATGCT 2780  
 1017 uLySgLyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAspL 1034  
 |||  
 2781 GAAAGGCTTGGGAGACATGTTCAAGTTTGGCAACATGCAAAAGATGACA 2830  
 1034 yslleGlyuLystrGlyLysIleLysIleGlnGlyuSerPheThrSerGly 1050  
 |||  
 2831 AGATTGAGAAACGGGTAAATTAATTAATTAATTAATTAATTAATTAATTA 2880  
 1051 GluGluArgIleArgMetLysGlnGlnGluArgIleGlnAlaLysTh 1067  
 |||  
 2881 GAGAGAGAGATGAGATGAGAGAGAGAGAGAGAGATTCAGACCCAAAC 2930  
 1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyraIleGlu 1084  
 |||  
 2931 TCGAGAAATTAAGGAGACAGACACTGAGAGCTGATGCTGAAATTC 2980  
 1084 lAspPheHisArgThrPheGlyCyAspAspGlyuLeuMetTyrglyGly 1100  
 |||  
 2981 AAGATTTTCAATCGGACATTTGGCTGTGATGATGATGATGATGATGATG 3030  
 1101 ValSerSerTyrglyGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117  
 |||  
 3031 GTTCTCTTATGAAAGTTCCATGCGCTCAACGCTGACCTCAGAGCC 3080  
 1117 oArgGluGlyHisMetLysAlaLeuTyraIleGlnValLysLysPro 1134  
 |||  
 3081 ACGAAGAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 3130  
 1134 rGAsnSerLySerProValAspSerAsnArgSerThrProSerAsn 1150  
 |||

3131 GGAATTCACAAACCTCACCTGTAGACAGTAAACAGTCAACTCCAGCAAT 3180  
 1151 HisAspArgIleGlnArgLeuArgGlnIlePheGlnIleAlaIysGlnAs 1167  
 3181 CATGATTCGATACAGCGCTGTAGCAGCAAGATTTCAGCAACCAAGCAAGA 3220  
 1167 pGluAspValGluAspArgArgArgThrTyrSerPheGlnGlnProTrp 1184  
 3231 TGAAGATGTGAAGATCGTGGCGACCTATGATTGTTGACCAACCCCTGGC 3280  
 1184 roAsnIleAspProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200  
 3281 CGAAGCAGACGGCCGGCCGACGACGCGGCGACACTCGGTGTCCTGGAG 3330  
 1201 ValGlnMetGlnArgGlnArgGlnGlnGluArgGlnSerSerGlnGlnAl 1217  
 3331 GTCCAGATGACAGCGGCGACGCGAGAGAGACCCGACAGCTCCCGACAGC 3380  
 1217 aGlnArgGlnTyrSerSerLeuProArgGlnSerArgLysAsnAlaSerS 1234  
 3381 CCAGCCGCACTACAGCTCTCTGCTCCGCAAGCAGCAAAATGCCAGCT 3430  
 1234 eValSerGlnAspSerTrpGlnIleAspTyrSerProGlyGlnGlyPhe 1250  
 3431 CGGCTCTCCAGACTCTTGGAGCAGAACTACTCCCTGGGGAAAGCTTC 3480  
 1251 GlnSerAlaIysGlnAsnProArgTyrSerSerTyrGlnIleSerArgAs 1267  
 3481 CAGAGTCCCAAGAAAGAACCCAGGTACTCCAGCTACCAAGGCTCCAGGAA 3530  
 1267 ngIlyTyrLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGluThrG 1284  
 3531 CGGCTACCTGGAGACATGCTTCAACGCCAGGCTCATCTGGAAACCTC 3580  
 1284 InGluLeuLeuArgGlnGlnGlnArgTyrGlyGlnGlnIleMetLysLys 1300  
 3581 AGGAGCTCTTCGCCAGGAAACAGAGCGGGAAGGAGCGCAGATACAGAG 3630  
 1301 GlnProProSerGlnGlyProSerAsnTyrAspSerTyrLysLysValG 1317  
 3631 CAGCTCTCTCCGAGGGGCCAGCAACTATGACCTGATATAAGAAAGTCCA 3680  
 1317 naSPProSerTyrAlaProProLysGlyProPheArgGlnAspValProp 1334  
 3681 GGACCCAGTACGCCCTCCCAAGGGCCCTTCGGCAGAGATGTCGCC 3730  
 1334 roSerProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGluLys 1350  
 3731 CCTCCCTCTCAGGTTCGAGGCTGAAACAGACTTCCAGACTCCTGAGGAA 3780  
 1351 GlyArgProPheTyrSer 1356  
 3781 GGGAGGCCCTTCTATTCC 3798

seq\_name: gb-pr-AF467005

seq\_documentation\_block: 3735 bp mRNA linear PRI 07-FEB-2002

LOCUS AF467005 3735 bp mRNA linear PRI 07-FEB-2002

DEFINITION Homo sapiens partitioning-defective 3 protein splice variant e (PARD3) mRNA, complete cds, alternatively spliced.

ACCESSION AF467005

VERSION AF467005.1 GI:18568349

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3735)

AUTHORS Gao, L., Macara, I. G. and Joberty, G.

TITLE Multiple splice variants of PARD3 and of a novel related gene, PAR3L, produce functionally different proteins

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3735)

AUTHORS Gao, L., Macara, I. G. and Joberty, G.

TITLE Direct Submission

JOURNAL Submitted (10-JAN-2002) Center for Cell Signaling, University of Virginia, PO Box 800577, Charlottesville, VA 22908, USA

FEATURES

source 1..3735

gene /organism="Homo sapiens"

1..3735 /db\_xref="taxon:9606"

1..3735 /gene="PARD3"

1..3735 /note="PAR3"

1..3735 /gene="PARD3"

/note="PAR3E: multi-PDZ domain containing protein; alternatively spliced"

/codon\_start=1

/product="partitioning-defective 3 protein splice variant e"

/protein\_id="AL76045.1"

/db\_xref="GI:18568350"

/translation="MKVVCFGRTRVVPCGDGMMKVFSLIOQAVTRYRKALAKDPNYWIOVHLEHGDGILDDILCDVADDDKRLVAVPEODPHHGDSSTGTSPEIGSELGNTNVSAFOPYOATSELEVPSYLRAAMPALVRSSDPALIGLSTSDSNFSSBPSRKNPTRMSTTAGFLKONTAGSPRTCDKDEDGTEEDRSRYEPGHADTGLEHINFSLDDVVKLVEVPNDGPGIIVHVPVSARGRLGLVRLKLEGGVAHEENLEFENDCIYRINDGILNRREFOAHMFROAMRTPITIMFHVPAANKEOYEOLOSSEKNMYSSRFSPSOYIDNRSVNSAGLHTVOAPRLNHPPEQIDSHSLPHSAHSPKPSAPA SAPONVESTTVSSGYTKKIGKRLNTOLKRGEGLEFITSRPTVITGSAPIYVKMLPGCAIADGRIRKAGDRLIVNGLVDYILKSOEFVYSLRSTKMGTYSLLVPEQEDAFHPREKRAEDDVIUTPDGTREFLEFEPPLNDSSAGAGVSKGRSKENHADGIFPKSILNGASADGRRLRVNDOLIVNGESLKTINDADNETLRSSTGKRNAGMLQIVARIRSKCNELKSPSPGPELPLETALDDDERIRSHLSYGLIGLIDSPSRNALSHIMGRYOLSPVNMQDDTVIITEDRLPVLPPLSDQSSSSSHSDDFEYTAGAGTAAKAIISDADCSLSPDVPVLAFOREGFRQSMSEKRTKOPSDASOLDFVTRKSKSDGSSPSHDVPSPLGIKSSSLSLQTAFAVETLNGDIPFHRPPRIIRGHCNESPFAAIDKSYDKPAVDDDEKMETLEEDTESSESSRESYSTASDPSHSLEROMNGOEGDXTDKDKTKGKKRKDKDREKDMKAKKMLGIGDMETIOAKTRFERROAKERYAETQDEHRTGCDDELMTYGGVSSYEGSMALNARPOSRECHMDALYAQVKPRNSPEVD SNTSTSNHDIRLRLROEQAKODEVEDRRRTYFEQPMVPAARATOSGRASVAYE VMQROROERERESQOAROYSLPROSKRNASVSODSMEQOYSPGEGPSAKEMPR VSSYOGSRNGYLGCHGFMAKRVMLFETQELLROBRREKQMKRPSBEGSNTDSYKVKQDPSTAPPKGPRFDVPPSPSOVARLNRLQTPEKGRFTS"

misc\_feature 682..942

/gene="PARD3"

/note="Region: PDZ domain"

1246..1506

/gene="PARD3"

/note="Region: PDZ domain"

1594..1851

/gene="PARD3"

/note="Region: PDZ domain"

misc\_feature

BASE COUNT 1090 a 860 c 990 g 795 t

ORIGIN

alignment\_scores:

Quality: 6342.00 Length: 1356

Ratio: 5.098 Gaps: 5

Percent Similarity: 91.740 Percent Identity: 91.667

alignment\_block:

US-09-757-781-2 x AF467005 ..

Align seg 1/1 to: AF467005 from: 1 to: 3735

1 MetLysValIleThrValCysPheGlyArgThrArgValAlaValProGlyG 17

1 ATGAAAGTACCGGTGCTTGGAGGACCCGGGGTGGTGGCCCTGGCG 50

17 YASPGLYHISMetLysValPheSerLeuIleGlnIleAlaValThrArgT 34

51 GGACGGCCACATGAAGATTTCAGGCTCATCCAGCAGCGGTGACCCGCT 100



1730 GAAGGCTTCGGTGAATGATCACTGATGACCTAAATGAGAAATCCCTG 1779  
 651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgSerMetSe 667  
 1780 TTGGGCAAGCAAAACCAAGATGCCATGGAAACCTTAAGAAAGGTCTATGTC 1829  
 667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684  
 1830 TACTGAAGGCAATAAACGAGCAATGATCCAGCTTAATGTTGCAAGGAA 1879  
 684 IeSerLysCysAsnGluLeuLysSerProGlySerProGlyProGlu 700  
 1880 TAAGCAAGTCAATGAGCTCAAGTCACTGAGAGCCCTGGAGCTGAG 1929  
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717  
 1930 CTGCCCATTTGAACAGCGGTTGGATGATAGAGAAAGAAATTTCCCATTC 1979  
 717 rLeuLysSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734  
 1980 CCTCTACAGTGGGATGAGGGGCTGTGATGAAATCCCCAGCAAAATGCTG 2029  
 734 IeLeuSerArgIleMetGlyLysSerGlyLysThrGlnLeuSerProThr 750  
 2030 CCTCAGTAGGATTAATG.....GGTAAATACCAAGTGTCCCTACA 2070  
 751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767  
 2071 GTGAATATGCCCCCAAGATGCACACTGATATATAGAAAGATGACAGGTGCC 2120  
 767 oValLeuProProHleLeuSerAspGlnSerSerSerSerHisAspA 784  
 2121 AGTGTTCCTCCACATCTCTACCAAGTCTCTCCAGCTCCCATGATG 2170  
 784 sValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800  
 2171 ATGTGGGGTTTGTGACGGCAGATGCTGTGCTGGGCCAAGGCTGCATTC 2220  
 801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817  
 2221 AGGATTTACACCGACGCTCTTGTAGTCCAGATGTTGATCCAGTCTTGGC 2270  
 817 aPheGlnArgGlyGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834  
 2271 TTTTTCACAGGAAGATTGTGAGCTCAGACTATCTGAGAAAAACGCCACAA 2320  
 834 ySgInPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850  
 2321 AGCAATTTTACATGCCAGTCAATTTGATTTCTGTTAAACACGAAATCA 2370  
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867  
 2371 AAAAGCATGGATTAGT..... 2388  
 867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884  
 2389 .....AGTCTCCACAGACAGATGTGGTCTTCCTGGGCTC 2425  
 884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGlyVal 900  
 2426 TGAAGAAGTCAAGCTCTGTGAGAAGTGTGCAGACCGCATGGCCGAGTGTG 2475  
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleArg 917  
 2476 ACTTTGAATGGGATATTTCTTTCATATGTCACAGCGCCGCGGAAATACAG 2525  
 917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerYrA 934  
 2526 AGCGAGGGGATGCAATAGAGATTCAGAGCTGCCATCGACAAAATCTTATG 2575  
 934 sPlysProAlaValAspAspAspAspGluGlyMetGluThrLeuGluGlu 950  
 2576 ATAAACCCGGGTAGATGATGATGATGAAAGCATGAGACCTTGGAAAGA 2625

951 AspThrGluLysSerSerArgSerGlyArgGluSerValSerThrAlaSe 967  
 2626 GACACAGAAAGAAAGTTCAAGATCAGGAGAGAGCTGTATTCACACAGCAG 2675  
 967 rAspGlnProSerHisSerLeuGluArgGlnMetAsnGlyAsnGlnGluL 984  
 2676 TGATCACCTTCCACCTCTGTGGAGACAAATGAAATGAAACCAAGAGA 2725  
 984 ySgLysAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000  
 2726 AAGGTGATTAAGACTGATAGAAAAAGATAAACTGAAAAAGAAAAGAG 2775  
 1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017  
 2776 AAGATAGAGATTAAGAGAAAGATAAATGAAACCAAGAAAGGAAATGCT 2825  
 1017 uLysGlyLeuGlyLysPheMetPheArgPheGlyLysHisArgLysAspL 1034  
 2826 GAAGGCTTGGAGACATGTTCC..... 2847  
 1034 ySIIleGluLysThrGlyLysIleLysIleGlnGlnLysSerPheThrSerGlu 1050  
 2847 ..... 2847  
 1051 GluGluArgIleArgMetLysGlnGlnGlnLysArgIleGlnAlaLysTh 1067  
 2848 .....AGATTCAAAGCCAAAC 2864  
 1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTrpAlaGluIleG 1084  
 2865 TCGAAGATTTAGAGGACGACAAAGCTCGAGAGCGTACCTATGCTGAAATTC 2914  
 1084 IAspPheHisArgThrPheGlyCysAspAspGluLeuMetYrGly 1100  
 2915 AAGATTTTCATCGGACATTTGGCTGTGATGATGATTAATATAGGGGGA 2964  
 1101 ValSerSerTyrgLysArgSerMetAlaLeuAsnAlaArgProGlnSerPr 1117  
 2965 GTTCTCTTATGAAAGGTTCCATGCTCTCAAGCTAGACCTCAGAGGCC 3014  
 1117 oArgGluGlyHisMetMetAspAlaLeuTrpAlaGlnValLysLysProA 1134  
 3015 ACGGAAGGGCATATGATGATGCTTTGTATGCCCAAGTCAAGAACGCCG 3064  
 1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150  
 3065 GGAATTTCCAAACCTCACCCTGTAGACAGTAAACAGATCAACTCCTAGCAAT 3114  
 1151 HisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAs 1167  
 3115 CATGATCGATACAGCGCTGTGAGCAAGAAATTTACAGCAAGCAAGCAAGA 3164  
 1167 pGluAspValGluAspArgArgArgArgThrTyrgSerPheGluGlnProTrpP 1184  
 3165 TGAAGATGTAGAAAGATCTGCGCGACCTATATGTTTGAAGAACCTTGGC 3214  
 1184 rOAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200  
 3215 CGAAGCGACGGCGCGGACGACGAGAGCGGCGCACTCGGTTCGTGGAG 3264  
 1201 ValGlnMetGlnArgGlnArgGlnGlnGluLysArgLysSerGlnGlnAl 1217  
 3265 GTGCAGATGCGACGGCGACGCGCAGAGAGAGCGCAAGACCTCCAGACGCG 3314  
 1217 aGlnArgGluTrpSerSerSerLeuProArgGlnSerArgLysAsnAlaSerS 1234  
 3315 CCAGCGCCAGTACAGCTCTGTGCTCGGCAAAAGAGAAATATCCACACT 3364  
 1234 eValSerGlnAspSerTrpGluGlnAsnTyrgSerProGlyGluGlyPhe 1250  
 3365 CGGTCTCCAGAGACTTGTGGAGACGAAGAACTACTCCCTGGGGAAGGCTTC 3414

1251 GlnSerAlaIysGluAsnProArgTyrSerSerTyrGlnGlySerArgAs 1267  
 |||  
 3415 CAGAGTGCACAAAGAGACCCCGAGTACTCCAGTACCAAGGCTCCAGAGAA 3464  
 1267 nclYTrLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGluThrG 1284  
 |||  
 3465 CGGCTACCTCGGAGGACATGCTTCAACGCCAGGCTCATGCTCGAGAACTC 3514  
 1284 InGluLeuLeuArgGlnGlnGlnArgTyrGlyGlnGlnMetLysLys 1300  
 |||  
 3515 AGAGACTCTCTCCCGCAGAAACAGAGCGGAGACAGACAGATGTAAGAAG 3564  
 1301 GlnProProSerGlnGlyProSerAsnTyrAspSerTyrLysLysValG 1317  
 |||  
 3565 CAGCTCTCTCCGAGGGCCACCAACTATGATCGCTATTAAGAAAGTCCA 3614  
 1317 nAspProSerTyrAlaProProLysGlyProPheArgGlnAspValPro 1334  
 |||  
 3615 GGACCCCAAGTACGCCCTCCCAAGGGCCCTCCGCAAGATGTCGCC 3664  
 1334 roSerProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGluLys 1350  
 |||  
 3665 CCTCCCTCTCAGAGTGCAGGCTGTAACAGACTCAGACTCTGAGAAA 3714  
 1351 GlyArgProPheTyrSer 1356  
 |||  
 3715 GGGAGGCCCTCTCTATTC 3732

seq\_name: gb\_to:AB005549

seq\_documentation\_block: 5500 bp mRNA linear ROD 13-NOV-1998  
 LOCUS AB005549  
 DEFINITION Rattus norvegicus mRNA for atypical PKC specific binding protein,

complete cds.  
 AB005549.1 GI:3868777  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE  
 1 (bases 1 to 5500)

AUTHORS  
 TITLE  
 JOURNAL  
 Iizumi, Y.  
 Direct Submission  
 Submitted (03-JUL-1997) Yasushi Iizumi, Yokohama City University  
 School of Medicine, Molecular Biology, 3-9, Fuku-ura, Kanazawa-Ku,  
 Yokohama, Kanagawa 236, Japan (E-mail: iizumi@med.yokohama-cu.ac.jp,  
 Tel:045-787-2597, Fax:045-785-4140)

REFERENCE  
 2 (sites)  
 Iizumi, Y., Hirose, T., Tamai, Y., Hirai, S., Nagashima, Y., Fujimoto, T.,  
 Tabuse, Y., Kempnes, K.J. and Ohno, S.  
 An atypical PKC directly associates and colocalizes at the  
 epithelial tight junction with ASIP, a mammalian homologue of  
 Caenorhabditis elegans polarity protein PAR-3

JOURNAL  
 MEDLINE  
 J. Cell Biol. 143 (1), 95-106 (1998)  
 98437350

FEATURES  
 source  
 Location/Qualifiers  
 1..5500  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /cell\_line="3Y1"  
 260..4273  
 /gene="asbp"  
 260..4273  
 /gene="asbp"  
 /product="atypical PKC specific binding protein"  
 /protein\_id="BA34216.1"  
 /db\_xref="GI:3868778"  
 /translation="MKVTVCFGRTRVVVPCGDKRKVSLIOATRYRKAVAPDPY  
 WIOVHLRHDGGIDLDIDILCDVADDKDRILAVAEQDPHGGGDTASSTGQSP  
 IPSELTGNNVSAFRPYQTSTSEIEVTPVSLRANMPLHVRSSDPALTLSTSVSDNNP

SSSEPRKNPTRWSTAGFLKONTTGSPTCDRKKENYRSILPRDSSWSNOFORNA  
 RSSLSHPEVNDMLEKROBDEGHEEDSRRVEPYGHADTGLENNPNSILDDMKLVQ  
 VPNDGGPGLIHVPSARGRTGLYKRLKFGKABDENFHEMDICIVRIINDDLRN  
 RPEQAOHMRQMRARVWTFVVPANKQYEQLOSOEMNNYSRGRSPSHCVANR  
 SVANNAFOALPRAPRLSOPPELDADHPLPHSAHASTRPPRLAPAPNVISTGVS  
 YNTRKGRRLNIOLKKTGELGFSITSRVDTIGGSAPLYVKNILDRCAIODGRKAC  
 DRILEVNGVDLACKSOEYVSLRSTKMGVSLLVFROEAFHFRKNAEESQMOBP  
 KERKAEDEDIYLPDGRREFLFEVPLNDGSGAGLGVSKYGRSKKENHADLIPKSI  
 INNGASCKDRERLAVNGESLIGKANOEAETLRSMSTEGNKRGMQLIVAR  
 RISCKNLSRPSRPAPELPIETLEDRERRISHLSYSGIEGLDESPTNMAISITMG  
 ESRCQSLSPVNMHDDTVATIEDRLPYLPILSDOSSSHDDVGFIMTEAGTWAKA  
 TISDSADCSLSPVDVPLAQRGFRGRQSKSEKRTKQPSNASQIDLPVTKRSGMDLG  
 IADETKLNTVDQDQRAQSPNDVPSLKKSSLSLELQTVAVAPLNLNGIIPHRPR  
 IIRRGCSNEFRAIDKSYDKPVDVDDDECMETLEETESSRSGRESYSTSDPSY  
 SLEROMNGDEPKRDKAEKKDKAGKDKDKREKDKLAKKMLGKIDMTFRGHR  
 KDKMETWGRKTIODSFTSEDRYRMKEQEQRIQAKTREFRQAPERYALIODPHR  
 TFEGDDELITYGNSYDGCIALANARPSPREGHIAMPTLYKQVKKPSSRPGDSNRSTP  
 SNHRIQLRQEROQAQODDVEDVRRRTYRFSQSSSRPASQSGHRSVVEVOVRO  
 ROERESFOAQROYSSLPROSRKNASVSVDSEWQNYARDEGFOSAKEPNRYSYOG  
 SRNGYLIGHGHNARVMLETDLLRQDORRRKBOULKQPADGVRRPQDVPSPSOV  
 ARNLQTPERKRPFFYS"

BASE COUNT 1450 a 1449 c 1547 g 1054 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 6332.50 Length: 1356  
 Ratio: 4.860 Gaps: 2  
 Percent Similarity: 96.091 Percent Identity: 89.676  
 alignment\_block:  
 us-09-757-781-2 x AB005549 ..  
 Align seg 1/1 to: AB005549 from: 1 to: 5500

1 MetLysValThrValCysPheGlyArgThrArgValValProCysG 17  
 |||  
 260 ATGAAAGTACCGCTGCTGGGAGGACCGGGTGGCGTGGCCGCGG 309  
 17 yAspGlyHisMetLysValPheSerLeuIleGlnAlaValThrArgT 34  
 |||  
 310 AGATGGCCCGCATGAAGATTTCAGCTTATCCAGCAGCGCGGTGAC 359  
 34 yArgLysAlaIleIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50  
 |||  
 360 ACCGGAAGCCCGTGGCAAGATCCAACTAGTACATCAGTACATCCG 409  
 51 LeuGlnHisGlyAspGlyLysLeuAspLeuAspAspIleLeuCysAs 67  
 |||  
 410 TTGGAGCAGTGGGAGCGGTGGATTCAGCTGATGATGATCTCTGGA 459  
 67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnAspP 84  
 |||  
 460 CGTTTCCGATGACAAAGACAGACTGTAGCAGTGTGATGAGCGAGTC 509  
 84 roHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100  
 |||  
 510 CCCACATGAGGAGATGATGATGATGATGATGATGATGATGATGAT 559  
 101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG 117  
 |||  
 560 CCACAGATATTCGGCAGTGGGCGGCACCAAGATTTCCTCTTTCG 609  
 117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134  
 |||  
 610 GCCTTATCAACACCAAGTGAATTTAGTACGCTTCAGTCTCTCGGG 659  
 134 IaAsnMetProLeuHisValArgTyrSerSerAspProAlaLeuIleGly 150  
 |||  
 660 CAATATCTCTTTCAGTCCGAGGAGAGTATGACCAAGCTTAACTGAG 709  
 151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerArg 167  
 |||



710 CTATCCACTTCGTCAGTGTATTAACAACTTTCTCTCAGAGAGAGCCCTCCAG 759  
167 glycAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184  
|||||  
760 GAAAAACCCACCGCGTGGTCCAGCAGAGCTGGCTTCTCAACAGACAGA 809  
184 hAlaGlySerProLysThrCysAspArgLysAspGluAsnTyrArg 200  
|||||  
810 CCACTGGAAGTCCCAAAACCTCGACAGAGAAAGATGAAACTACAGC 859  
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217  
|||||  
860 AGCTCCACAGGAGTCCAGTACTGGTCCACCAAGTCCAGAGAGACAA 909  
217 hAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeu 234  
|||||  
910 CGCCCGCTCTCTGTAGTCCAGCCACCAATGTAAGACCGGTGGTGG 959  
234 LuLysGlnGluGlnAspGlyThrGluGluAspAsnSerArgVal 250  
|||||  
960 AGAAGCAGAAACGAGCAGAGAGAGCCACAGAACAGACACCGCAGT 1009  
251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPhe 267  
|||||  
1010 GAGCCGTTGGACATGCTGATCTGGCTGGAGATATGCCCCAATTTC 1059  
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284  
|||||  
1060 CCTCGATGATATGTAAAGCTCTACAACTCCCAACGATGGAGGGCCC 1109  
284 euGlyIleHisValValProPheSerAlaArgGlyLysArgThrLeuGly 300  
|||||  
1110 TGGGAATCCATGTAGTGCCTTTCAGTCTCGAGCGGCAGAACACTGGG 1159  
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisIleuAsnLe 317  
|||||  
1160 TTTTGTGTAAGCGGTTGGAAAGGTGGAGGCTGACCAAGAAACCT 1209  
317 uphArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334  
|||||  
1210 TTTCCAGCAGAAATGATGCTGATGTGAGGATTAAGATGGATCTTGCA 1259  
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350  
|||||  
1260 ACAGAAGATTGGAGCAAGCAGCATATGTTCCGCCCAACCATGCGTGC 1309  
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnTyrG 367  
|||||  
1310 CGAGTCTTTGGTCCAGCTGGTCCCTGCAGCCAGAACAGAGCAGTATGA 1359  
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384  
|||||  
1360 ACAGCTCTCCCAACGAGATGAACAATCTATCCCTGGCGGTCTCAGCC 1409  
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerIleGlyLeuHis 400  
|||||  
1410 CCACAGACCCACTGTGTGGCCACAGAGAGTGTGCCCAATGCGCCCTCAG 1459  
401 ThrValGlnArgLapProArgLeuAsnHisProProGluGlnIleAspSe 417  
|||||  
1460 GCGTTGCCAGAGACCCACAGCTGAGCCACACCTGAGCAGCTGGAGTGC 1509  
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProserA 434  
|||||  
1510 TCACCCCGGACTACCTATAGTGTCTATGCTCACAACCAACCGCCACAG 1559  
434 laProAlaSerAlaProGluAsnValPheSerThrThrValSerSerGly 450  
|||||  
1560 CCCACGCTTGGCTCCACCAATGTCTGTAGTACCACTGTGCGCAGTGTG 1609  
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467  
|||||  
1610 TACAAACGAAAGAGATGAGCAAGAGGCTCAACATTGACTTAAAAAAG 1659

467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484  
|||||  
1660 TACAGAAAGACTGGGATTTAGCATCACTTCCGAGAGCTACTATAGT 1709  
484 LysSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaIle 500  
|||||  
1710 GCTCAGTCTCCATTTATGTACAGAACATTTCTCTCCAGAGGGGCTCAT 1759  
501 GluAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517  
|||||  
1760 CAGATGCGACGCTCAAGCGGGAGACCGCTTATATAGGCTCAATGGAGT 1809  
517 laAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerT 534  
|||||  
1810 AGATTATGCGGCAAGTCCAGAGAGAAATGTTTCCCTGTGGAGAAACA 1859  
534 hTrpLysMetGluGlyThrValSerLeuLeuValPheArgGlnLysPala 550  
|||||  
1860 CCAAGATGAGGGGACCGTGAAGCTTGTGGTCTTTCGTCAAGAAAGAGCT 1909  
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567  
|||||  
1910 TTCCACCCAAAGGAAATGAATGCCGAACCCAGCCAGATGCCAGACTCCAA 1959  
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584  
|||||  
1960 AGAAACAAAGCCGAAACAGCAGACGATTTGTTCTCACACCTGACGATNCA 2009  
584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600  
|||||  
2010 GGGGTTTCTGACTTTCGAAGTTCTCACTGAATGACTCAGGGTCTGCAAGG 2059  
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617  
|||||  
2060 CTTCGTCTACGCTCAAGGGGACCGCTCCAAAGAAACCATGCGAGATT 2109  
617 uGlyIlePheValLysSerIleIleAsnGlyLysAlaLaserLysAspG 634  
|||||  
2110 GGGGATCTTGTTGAATGATCATCATTAACGGGGGAGCTGCATCTAAAGATG 2159  
634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysLeu 650  
|||||  
2160 GAGAGCTGAGAGGTAAATGACCAGCTGATAGCTGGAACGGAGACTCTGT 2209  
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667  
|||||  
2210 CTGGGCAAAAGCCAAACGAAAGCCATGGAGACTGTGAGAGAGTCAATGTC 2259  
667 rThrGluGlyAsnLysArgLysMetIleGlnLeuIleValAlaArgArg 684  
|||||  
2260 CACCGAGGCGCAACAGCGTGGCATGATCCAGCTCATGCTGGCGAGGGGA 2309  
684 leSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700  
|||||  
2310 TCACACAGATGAACGAGTTGGGTCTCTGTGGACCCCGCTCAGCCGAG 2359  
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717  
|||||  
2360 CTGCCCTAGAAACAGAGATTGGATGACAGAGAACGCGAGATCTACACTC 2409  
717 rLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734  
|||||  
2410 CCTTACAGTGGGATTCGAGGGGCTGAGTACATCTCCACACGAGATGCG 2459  
734 laLeuSerArgIleMetGlyLysSerGlyLysTyrGlnLeuSerProThr 750  
|||||  
2460 CACTCAGCAGAGATATATGGGTGAGTCAAGTAAATCCAGCTCTCCCAAC 2509  
751 ValAsnMetProGluAsnAspThrValIleIleGluAspAspArgLeuP 767  
|||||  
2510 GTGAACATGCCACATGATGACACTGTCTATGTGAAGATGACAGAGTGC 2559

767 oValLeuProProHisLeuSerAspGlnSerSerSerHisAspA 784  
 |||||  
 2560 TGTGCTCCCTCTCACTCTGTGACCAAGTCTCTCCAGCTCCATGATG 2609  
 784 spValGlyPheValThrAlaAspAlaGlyThrTriPAlaLysAlaIaIe 800  
 |||||  
 2610 ACCTGGGATTCATTAATGACAGAGCGAGCTGGGCCAAGCTTACCATC 2659  
 801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817  
 |||||  
 2660 AGTGATCTACCGGACTGCTCATTTGAGTCCAGATGTTGATCCGTTCTCCG 2709  
 817 apheGlnArgGluGlyPheGlyArgGlnSerMetSerGlyLysArgThrL 834  
 |||||  
 2710 TTTTCAACAGAGAGGATTTGGACGCCAGATGTGTAGAAAAACGACAA 2759  
 834 yslGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850  
 |||||  
 2760 AGCAATTTTCAATGCCAGTCAATTTGATTTGCTTAAACAGCAAAATCA 2809  
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867  
 |||||  
 2810 AAAAGCATGATTTAGTATAGCTGACGAGACTMAACTCATATCACTGGA 2859  
 867 paAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884  
 |||||  
 2860 TGACCGAGAGCGAGGCTCCCCCAATAGATGTGGAGCCCTCTTGCGTCT 2909  
 884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900  
 |||||  
 2910 TGAAGAAATCCAGCTCTTGTAGAAAGTCTGCAGACGGCTTGTCTGAGTCT 2959  
 901 ThrLeuAsnGlnLysPheProPheHisArgProArgProArgIleIleArg 917  
 |||||  
 2960 ACCCTGATGTGGAAACATTTCTTCCACCGCCACGGCCAGAAATCATCCG 3009  
 917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerTyrA 934  
 |||||  
 3010 AGGAGAGGGCTGCACAGAGAGCTTCAGAGCCGCTTGTGACAAATGCTTACG 3059  
 934 spLysProAlaValAlaAspAspAspGlnGlyMetGluThrLeuGluGlu 950  
 |||||  
 3060 ATAGCCCATGTGTGATGACGACAGCAAGCATGAGACCTTGGAGAA 3109  
 951 AspThrGlnGlnSerSerArgSerGlyArgGlnSerValSerThrAlaSer 967  
 |||||  
 3110 GACACAGAAAGAAAGTCAAGGTCAGGAGGAGGAGCTCCTGTCCAGTCCAG 3159  
 967 rAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlyAsnGlnGlu 984  
 |||||  
 3160 TGATCAGGCTTCTTAATCTCTGAGAGCAAAATGAAATGAGAACCCAGAGA 3209  
 984 yslGlyAspLysThrAspArgGlyLysAspLysThrGlyLysGluLysLys 1000  
 |||||  
 3210 AGAGGAGAACAGAGAGAGAAAGAAAGCAAAAGCCGGAAAGGATTAAGAG 3259  
 1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017  
 |||||  
 3260 AAAAGCCAGAGAGAGAGAAAGATTAATGAAAGCCAAAGAGGGAGATCT 3309  
 1017 ulysGlyLeuGlnLysPheMetPheArgPheGlyLysHisArgLysAspL 1034  
 |||||  
 3310 GAAAGGCTTGGGGGACATGTTCAAGTTTGGCAAAATCGAAAGAGTGCACA 3359  
 1034 yslIleGlnLysThrGlyLysLysLysIleGlnGlnLysPheThrSerGlu 1050  
 |||||  
 3360 AGATGGAAGAAATGGTCTGAATTAATTAATCCAGACTCTTTCACTCAGAA 3409  
 1051 GluGluArgGlyLeuArgMetLysGlnGlnGlnLysArgIleGlnAlaLysTh 1067  
 |||||  
 3410 GAGGACAGAGGCTCGGATGAGAGAGAGACAGAGAGATTTCAAGCCAAAC 3459  
 1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyrIleGluIleG 1084

|||||  
 3460 TCGAAGATTTTAGGAGCGGCAAGCTCGAGAGCCGACTATGCTGAATTC 3509  
 1084 lnaAspPheHisArgThrPheGlyCysAspAspGluLeuMetTyrGlyGly 1100  
 |||||  
 3510 AAGATTTCCATCGAACAATTTGGCTGTGTATGATGAGAGTTGCTGTATGGGGG 3559  
 1101 ValSerSerTyrGlnGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117  
 |||||  
 3560 ATGTATCTCTTATGATGGCTGGCTGCTCTCAATGACAGACCCAGAGCC 3609  
 1117 oArgGlnGlyHisMetLeuAlaLeuTyrAlaGlnValLysLysProA 1134  
 |||||  
 3610 AAGACAAAGTCACTCATGATGACACTTTGTATGACAAAGTAAAGAAACCTC 3659  
 1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150  
 |||||  
 3660 GGAGCTCCAAACCTGGG.....GACACCAATTCATCCACCCTTAGACAC 3703  
 1151 HisAspArgIleGlnArgLeuArgGlnGlnPheGlnGlnAlaLysGlnAs 1167  
 |||||  
 3704 CATGACCGGATACAAAGCTTACGGCAGCAAGATTCACAGAACCAAGAGA 3753  
 1167 pGluAspValGluAspArgArgArgThrTyrSerPheGlnGlnProTrpP 1184  
 |||||  
 3754 TGAGACGCTGGAAGACGGCGCGTACCTACAGCTTGAAGCAATCTTGTCT 3803  
 1184 roAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200  
 |||||  
 3804 CCAGCTCAAGGCGCACATCGCAGAGTGTGTGCGGCTGCACTGCACTGAG 3853  
 1201 ValGlnMetGlnArgGlnArgGlnGlnGluGluGlnLysSerGlnGlnAl 1217  
 |||||  
 3854 GTTCAAGTGCAGCGGCGAGCGCCAGAGAGGAGCGAGAGCTTCCACAGGC 3903  
 1217 aGlnArgGlnLysSerSerLeuProArgGlnSerArgLysAsnAlaSerS 1234  
 |||||  
 3904 CCAAGCCGCAATACAGTCTCCTCTCCAAAGACAAAGACGAAATGCCAGCT 3953  
 1234 erValSerGlnAspSerTrpGlnGlnAsnTyrSerProGlyGluGlyPhe 1250  
 |||||  
 3954 CTGTATCACAGATTCCTGTGGAACAGAACTACGCCCTGTGAAAGCTTTC 4003  
 1251 GlnSerAlaLysGlnAsnProArgTyrSerSerTyrGlnGlySerArgAs 1267  
 |||||  
 4004 CAGAGTCCAGAGAGAAACCCAGGATTTCCAGTTCACAGGCTCCAGGAA 4053  
 1267 nGlyTyrLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGluThrG 1284  
 |||||  
 4054 CGGCTATCTTAGTGGCCATGGCTTTAATGCAAGGTCATGCTGAGACACC 4103  
 1284 lGlnLeuLeuArgGlnGlnGlnArgArgLysGlnGlnGlnMetLysLys 1300  
 |||||  
 4104 AGGAGCTTCCCGACAGAGAGCAGAGCCGAAAGACAGACAGACGAGAGAG 4153  
 1301 GlnProProSerGlnGlyProSerAsnTyrAspSerTyrLysValGlu 1317  
 |||||  
 4154 CAGCCACCAAGGAGTGA..... 4171  
 1317 naSProSerTyrAlaProProLysGlyProPheArgGlnAspValProp 1334  
 |||||  
 4172 .....GTCAAGAGGGCCCTTCCGGCAAGATGATCCGCC 4202  
 1334 roSerProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGluLys 1350  
 |||||  
 4203 CGTCTCATCTCAGGTGGCCAGGCTGAACAGACTGCAGACACCCGAGAAA 4252  
 1351 GlyArgProPheTyrSer 1356  
 |||||  
 4253 GGGCGGCTTCTTACTCTC 4270  
 seq\_name: gb\_ro:Ay026057

seq\_documentation\_block:  
LOCUS AY026057 4167 bp mRNA linear ROD 20-FEB-2001  
DEFINITION Mus musculus PAR-3 180 kDa isoform (Par3) mRNA, complete cds.  
ACCESSION AY026057  
VERSION AY026057.1 GI:12965348  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS 1 (bases 1 to 4167)  
TITLE Lin, D., Gish, G.D., Songyang, Z. and Pawson, T.  
The carboxyl terminus of B class ephrins constitutes a p02 domain  
binding motif  
JOURNAL J. Biol. Chem. 274 (6), 3726-3733 (1999)  
MEDLINE 99121117  
PUBMED 9920925  
REFERENCE 2 (bases 1 to 4167)  
AUTHORS Lin, D., Edwards, A.S., Fawcett, J.P., Mbanalu, G., Scott, J.D. and  
Pawson, T.  
TITLE A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and APC  
signalling and cell polarity  
JOURNAL Nat. Cell Biol. 2 (6), 540-547 (2000)  
MEDLINE 20394297  
PUBMED 10934475  
REFERENCE 3 (bases 1 to 4167)  
AUTHORS Lin, D.C. Jr. and Pawson, T.J.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-2001) Programme in Molecular Biology, Samuel  
Lunenfeld Research Institute, 600 University Avenue, Toronto,  
Ontario M5G 1X5, Canada  
FEATURES  
source Location/Qualifiers  
1..4167  
/organism="Mus musculus"  
/strain="NIH Swiss"  
/db\_xref="taxon:10090"  
1..4167  
/gene="Par3"  
110..4111  
/gene="Par3"  
/note="nPAR-3: scaffolding protein"  
/codon\_start=1  
/product="PAR-3 180 kDa isoform"  
/protein\_id="AAK07669.1"  
/db\_xref="GI:12965348"  
/translation="MKTVVCEGRTRVVVPCGGRMKVPSLIQOAVTRYRKAVAKDPN  
WIOYRLRHGGGILDDLDICDVADDDKRLVAVEDBDPHGGGCTASPTGQSP  
IFGSEICNNYSARQPYOATSEIETVPSYLRAMPVLRHRRSDPLITGLSTSYNNF  
SSEPSRKNPTRMSTTAGFLKONTAGSPRTCDRKDEYRSLPRPSSNSQFDNA  
RSSLSASHPNVDRMLEKODEEGTEEDSRYEPVGHADYLENNPNSLIDMVKLYO  
VPNDGPIGLIHVPFSGARGRTIGILVRLKRGKAEQENLFHENDCIVRIINDGLRN  
RRFEQAHMFRQAMRAVIMFHVYPAKKEQEQSREKKNYSGGRPSDGHCVANR  
SVANNAPOALPRAPRLSQPQLDAPRLPHSAHASTKPPAPALAPSVYSTNGSV  
YNTKKVGRRLTIQKKKEGIGFSTSDYITGSAPIYVKNILREGAIIQDGRKAG  
DRLTEVNVDLAGKSQEEVSLRSTKMEGVSLVTFQEEAFHRENNAEBSQMQPT  
KETAEDEVDVLTGDTRELFTEVPLDGSAGVGAVGKRSKRENAADLIEFKSI  
INGGAASKDGLRVNDQLIAVNGESILGKANOAEETLRSGSTGNKRNQMLIVAR  
RISRCNELRSPSGPAPELPIETELDDERRRSHELYSIEGLDSPRNALSLRMG  
KCOLSPYNNMHDVTMIEDRLPYLPHLISDSSSSSHDDYGFIMTEAGTMAKRTIS  
DSACSTSPDYVPIVAFORREGRGKSMSEKRTKQSDASQDLDFYTRKSKMDLYADE  
TKLTVDDQKRGSPRDVGPISGLKSSLESLOTVAVELVTLNGNIPHRPRLIRG  
KGNESFRAAIDKSTDKPMVDDDEGMETLEDEESSRGSSEYSTSDPSYLER  
OMNGDPERKDRIDKTKKADKDKKREKEDKLANKGMLKGLDMRFKGRDCK  
MEKMRIRIKIODEFTSEEDRVNMEKEQERLOATREFRERARERDYAEIODEHFFGC  
DDELLIGMSSYEGCLALNARPOSPREGLMDTLYAOKKPPSSKPGDSNRTENHD  
RIORROFOQAKODEVEDERRRYSPEOSWSSSPASQSGHSYSEVYQVORQOE  
RESTRQARQYSSLPKRSKNAASSISQSWENOTAPGGEGFSAKNPRTSSQSRNG  
YLGGEFNARVWLETQELLKQDQRKEQOLAKOPPADVGRGFQDVPSPQVARN  
RLQPEKRRPYS"

BASE COUNT 1145 a 1079 c 1174 g 769 t  
ORIGIN

alignment\_scores:  
Quality: 6310.50 Length: 1356  
Ratio: 4.858 Gaps: 4  
Percent Similarity: 95.796 Percent Identity: 89.676  
alignment\_block:  
US-09-757-781-2 x AY026057 ..  
Align seg 1/1 to: AY026057 from: 1 to: 4167  
1 MetysValthrValCysphcGlyArgThrArgValValAlProCysGI 17  
110 ATGAAGTGAACCGTGTCTTGGAGAGACCGGGGTGGTGGCGTGGG 159  
17 YASPGIYHISmetysValPheSerLeuILEGIlnAlaValThrArgT 34  
160 AGATGCCCGCATGAAGTTTTCACCTTCATTCACAGCGCGGTGACCC 209  
34 YRArgLysAlaIleAlaLysAspProAsnTYRTPILeglnValHisArg 50  
210 ACCGGAAGCGCGTGGCCAAAGATCCAAACTGATGATACAGTGCATCGC 259  
51 LeuGIuHisGlyAspGlyGlyIleLeuAspLeuAspAspIleLeuGysAs 67  
260 TTGGAGCATGGAGATGGAGGATTTCTAGACCTGATGACATCCCTGTGA 309  
67 PValAlaAspAspLysAspArgLeuValAlaValPheAspGIuGlnAsp 84  
310 CGTTGCTGATGACAAAGACAGACTGTGATGACTTTTGATATACAGATC 359  
84 roHISHisGlyLysAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100  
360 CCCACCATGGAGAGATGTGTACAGCGGCACGCTTCACGGGAACCAAGT 409  
101 ProGIuIlePheGlySerGluLeuGlyThrTrsAsnValSerAlaPheGI 117  
410 CCAAGATATTGGGACGTGAGTGGGCAACAAATGTTCTGTGCTTTCA 459  
117 nProTYGlnAlaThrSerGIuILEGIlnAlaValThrProSerValLeuArg 134  
460 GCCTTATCAACCCCAAGTGAATTTGAAGTACACCCCTTCACTTCCGG 509  
134 IAsnMetProLeuHisValArgArgSerSerAspProAlaLeuILEGly 150  
510 CAATATGCGCTTCATGTCGCCGCGAGACAGCACCGCTTAACTGGC 559  
151 LeuSerThrSerValSerAspSerAsnPheserGluGluProSerArg 167  
560 CTTCACCTTCTGTCACTGATMAACAATTTTCTCAGAGGAGCCCTCCAG 609  
167 GlyAsnProThrArgTrpSerThrAlaGlyPheLeuLysGlnAsnT 184  
610 GAAAAACCCACCGCGTGTCCACAGACAGCGCTTCTCAACGAGAACA 659  
184 hralaGlySerProLysThrCysAspArgLysLysAspGIuAsnTYArg 200  
660 CCGGTGAAGTCCCAAAACCTGCGACAGAGAAGAAAGATGAACATACAGA 709  
201 SerLeuProArgAspThrSerAsnTrpSerAsnIlePheGlnArgAspAs 217  
710 AGCTTCCACGGGATCCACAGTGTGTCACCAAGTTCACGCGACACAA 759  
217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeuG 234  
760 CGCCGCTCCTCCTGTAGCCGACACCAATGATGATAGACCGGTGCTGG 809  
234 LuYsGIuGlnGluAspGlyThrGluGluAspAsnSerArgVal 250  
810 AGAAGCAAGAACAGATGAGGAAGACACAGAAAGAACAGACGCGAGTG 859  
251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPhese 267  
860 GAGCGGTGGACATGCTGATACCGGATGGAGAACATGCGCCAACTTTTC 909

```

267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyProL 284
910 CCTCGATGATATGTAAGCTGTCAAGTCCCAAGATGAGGGCCCC 959
284 euGlyIleHisValAlaProPheSerAlaArgGlyArgThrLeuGly 300
960 TGGGAATCATGTAGTGGCTTTCAGTCTCGAGCGCGCAACAATGGGG 1009
301 LeuLeuValLysArgLeuGlyGlyGlyLysAlaGluHisGluAsnLe 317
1010 TTGTAGTGAAGCGGTGGAGAAAGCGGTAAAGCTGAGCAAGAAACCT 1059
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
1060 TTTCATGAGAAATGACTGCAATGTGTGAGATTAAAGATGAGATCTGCA 1109
334 snArgArgPheGluGluAlaGluHisMetPheArgGluAlaMetArgThr 350
1110 ACAAGAAATTTGAGCAAGCAACAATATGTTCCCGCAAGCTATGCGCG 1159
351 ProIleIleTrpPheHisValAlaProAlaAlaAsnLysGluGlyArg 1209
1160 CGTGCTATTGGTTCATGTGTGCTCCGTCAGCAAAACAAGAGCAATATGA 1259
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrrTyrrSerArgPheSerP 384
1210 ACAACTGTCCCAAGCGGAGAAACAACACTACTCCCAAGCGCGCTTATGCC 1259
384 roAspSerGlnTyrrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
1260 CTGACAGCCACTGTGTGGCCACAGAGATGTGGCCCAACAATCCCCCTCA 1309
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
1310 GCATTGGCCAGACACCCAGACTGAGTCAAGCCACCGAGAGCGTGGATGC 1359
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProSerArg 434
1360 TCACCCCCGACTACCTCATAGTGTCTACGCTCAACCAACCAACCCGAG 1409
434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1410 CCCCGGCTTGGCTCCACCCAGTGTGTAGTACCAAGTAAAGGAGTGTG 1459
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysGly 467
1460 TCAACACAGAAAGAAAGTAAAGCAAGAGGCTCAACATCAGCTTAAGAAAG 1509
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484
1510 TACAGAAAGACTGGGATTTCAGCATCACCTCCGGGACGTCCACCAATAGTG 1559
484 LysSerAlaProIleTyrrValLysAsnIleLeuProArgGlyAlaIle 500
1560 GCTCAGCTCCCATTTATGTCAAGAAATATCTTCCTCGAGGGGCTGCCATT 1609
501 GluAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVal 517
1610 CAGAGTGGCAGACTCAAGCAGAGACCGGCTAATAGAGTCAATGAGCT 1659
517 laAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSer 534
1660 AGATTTAAGCAGCAATCCAGAGGAAAGTGTTCCTGTTGGAAGCA 1709
534 hLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
1710 CCAAGATGGAGGAGCTGTGAGCTCTGCTTCTTCAGGAAAGGCT 1759
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProL 567
1760 TTCCACCCCAAGGAATGATGCTGTAACCAAGCAGATCAGACATCTCAAA 1809

567 sGluThrLysAlaGluAspGluAspIleValIleLeuThrProAspGlyThr 584
1810 AGAAACGAAGGCTGAAGATGAGAGCTGTGTTCTCACACCCAGATGACA 1859
584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
1860 GGGAGTTTCTGACTTTTGAATCTTCAGTATGATCAGAGATCTGCAGGG 1909
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
1910 CTGTGTGCTAGTGTCAAGAGGAAACCTTCCAAAGAAACACCGAGATT 1959
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
1960 GGGGATCTTGTTAAATCCATTATCATATGATGAGACTCATCTAAAGATG 2009
634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysLeu 650
2010 GAAGGCTGAGCGTAAATGACAGCTGATACCTGTCAATGCAATCTCTTA 2059
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
2060 CTGGCAAAAGCCACACAGAAAGCCATGAGACTTACGAGGATTCATGTC 2109
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684
2110 CACCGAGGGCAACAAAGCTGCATGATTCAGCTCATTTGTGGAGGCGGA 2159
684 leSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGly 700
2160 TCAGCAGATGTAAAGAGCTCGGCTCTCTGAGCCCTCTCAGCTCAGTAG 7209
701 LeuProIleGluThrAlaLeuAspAspArgGluArgGlyIleSerHisSe 717
2210 CTGCCCATCGAGACAGAACTGTATGCCAGAACGAGATCTCACACTC 2259
717 rLeuTyrrSerGlyIleGluGlyLeuAspGlyLysSerProSerArgAsnAla 734
2260 CCTCTACAGTGGGATGAGAGGGCTGTGATCAGTCGCCACAGAAACGAG 2309
734 laLeuSerArgIleMetGlyGluSerGlyLysTyrrGlnLeuSerProThr 750
2310 CACTGAGCAGAGTAATG.....GCTAAATCCAGCTGTCCCCCAG 2350
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767
2351 GTGAACATGCTCAGATGACATGATCATGATGAAAGATGACAGGCTGCC 2400
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784
2401 TGTGCTCCCTCTCCTCCTGACAGCTCTCTCTCAGCTCCCATGATG 2450
784 sPValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800
2451 ACCTTGATTCATATGACAGAGCAGCAGCTGGGCCAAACCTACATC 2500
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValIleAl 817
2501 AGTACATCAGCCGACTGCTCATTTGATGTCAGATGTTGATCGGTTCTGC 2550
817 aPheGlnArgGlyGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
2551 TTTTCAACGGGAAGATTGGAGCGCAGAGATATCTCGAAAAAGCACA 2600
834 yGlnIlePheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
2601 AGCAATTTTCAGATCCAGTCAATTTGATTGCTTAAACACCAAAATATA 2650
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
2651 AAAAGCATGATTTA...GTAGTGAAGAGACTAAACTCAATACAGTGA 2697
867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884

```



## TITLE

Direct Submission

## JOURNAL

Submitted (18-Oct-1999) Laboratory of Molecular and Cellular  
Oncology, Shanghai Institutes for Biological Sciences, Chinese  
Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, P. R. China

## FEATURES

Location/Qualifiers

1..3470

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="10"

/map="10p11.2"

288..3383

/note="hasRpsa"

/product="atypical PKC isoform-specific interacting

protein short variant"

/protein\_id="AAK27892.1"

/db\_xref="GI:13491612"

/translation="MKYVCFGRTRVVPCGDSHMKVPSLIGQAVTRYRKAIADPNY

IIGSELGTNNVSAFOYQATSEIYVPSYLRANMPLHVRSSDPAIIGLSVGSNP

RSSEPSRNPTRMSTTAGFLKONTASPTCKDKDENYRSLPRTSMNSMOPRONA

VPNDGPGIHYVPEFSAKGRITLGLVLRLEKGRKAHEHNFRENDCTVRLNDKYLE

SYNSAGLHYVORARLNLHPEDQIDSHSLRPHSAHSGKPSAPADQNESTYDNR

YNRKIKGRNLQKLGTEGLGFSITSRDVIGTSAPRYKANILPRGAAIODERLKAG

DRLIEVNGDVLRKSOEYVSLRSTKMEGVSLIVRQEDAEHPRDLNAEPSONOIP

KETRAEDEDVLPDGRFELFEVPLNDSSAGLGVSVGNSKEVHMDLGFYVSI

RISKNASKGRRLRVNDQILAVNGESILCKTMDMELVRSKNSKEVHMDLGFYVSI

KYSKMEKSPGSPGPELPIETALDDREKISLSISLGLGEGENKRMQIOLYAR

KYQSLPTVMPPQDDTVIIEDRLVLPPLHSDSSHDVDFVATPDAADMAKAIIS

DKADSLSPDVPVLAFOREGFRKQMSSEKRTQFSPASQDLQVTKRKSKMDLQAD

ETRLNTVDDOKAGSPSRPGLSLKSSLSLELQTAVALVETLNDIPIPHRRPRILR

GRCNESFRALIDKSYDKPAVDDECEMTELEEDTESSRSRPSVSTASDPSISLE

ROMANGNOGKGTDRKKDKTEKKDKRDKKDKKAKKAKMLKGLGDMFSIAKLPEK

R"

BASE COUNT

1003 a

811 c

934 g

721 t

1 others

ORIGIN

alignment\_scores:  
 Quality: 5268.50  
 Ratio: 5.130  
 Percent Similarity: 99.323  
 Length: 1034  
 Gaps: 1  
 Percent Identity: 98.839

## alignment\_block:

US-09-757-781-2 x AF196186

Align seg 1/1 to: AF196186 from: 1 to: 3470

1 MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17  
 288 ATGAAAGTACCGTGTGTCGAGACGACCGGTCGTCGTCGCGG 337  
 17 yAspGlyHisMetLysValPheSerLeuLleGlnGlnAlaValThrArgT 34  
 338 GGAGGCGGCACATGAAGTTTCACCTTCACAGACGCGGTGACCCGCT 387  
 34 yAlrGlyAlaLleAlaLysAspProAsnYTrpLleGlnValHisArg 50  
 388 ACCGGAAGCCATCCGCAAGATCCAACTACTGATACAGGTGACATCC 437  
 51 LeuGlnHisGlyAspGlyLleLeuAspLeuAspAspLleLeuCysAs 67  
 438 TGGAACTGAGAGTGAAGAACTAGACCTTGATGACATCTTTGTGA 487  
 67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnGlnAsp 84  
 488 TGTAGCAGACGATAAAGACACCTGAGTACAGTGTGATGACAGATC 537  
 84 roHSHISGLYGLYAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100  
 538 CACATCAGCGAGGTGATGACACAGTGCAGTTCACCGGTACCCAGAGC 587

101 ProGluLlePheGlySerGluLeuGlyLysThrAsnAsnValSerAlaPheG1 117  
 588 CCAGACATATTGTTGTTAGTACGCTGGCACCAACAAATGTCACACCTTTTA 637  
 117 nProYrGlnAlaThrSerGluLleGluValAlaThrProSerValLeuArg 134  
 638 GCCTTACCAAGCAACAGTGAATTTGAGTGCACACCTTACGCTTCAG 687  
 134 lAsnMetProLeuHisValArgArgSerSerAspProAlaLeuLleGly 150  
 688 CAATATCCCTCTTCATGTTCCAGCCAGTGAAGTACCGCTCTAATTCGC 737  
 151 LeuSerThrSerValSerAspSerAsnPheSerSerSerGluGluProSerAr 167  
 738 CTCTCCACTTCTCTCAGTATAGTAAATTTTCCTCTGAAGACCTTCAAG 787  
 167 GlyAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184  
 788 GAATAATCCACACCGCTGTCMAACACGCTGCTTCTCAACACAGACA 837  
 184 hAlaGlySerProLysThrCysAspArgLysLysAspGluAsnYTrArg 200  
 838 CTGCTGGGAGTCTTAAACCTGCGACAGAAAGAAAGATGAAACTACAGA 887  
 201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217  
 888 AGCTTCCGCGGAGTACTAGTAACTGCTTAACCAATTTCAGAGACAA 937  
 217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234  
 938 TGCCTGCTCTCTCTGAGTGCACGTCACCAATGGTGGCAAGTGGCTGC 987  
 234 LuLysGlnGluAsnAspGlyLysArgGlyThrGluGluAspAspSerArgVal 250  
 988 AGAAACAGACAGATGAGATGGAGTGGACAGAAAGATTAACAGTCTGT 1037  
 251 GluProValGlyHisAlaAspThrGlyLeuGlnHisLleProAsnPheSe 267  
 1038 GAACCTGTGGACATGTCGACACGCGGTTTGACATATACCAACTTTTC 1087  
 267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284  
 1088 TCTGATGATATGTAAGCTCCGTAAGATCCCAACAGATGAGGCGCTC 1137  
 284 euGlyLleHisValAlaProPheSerAlaArgGlyGlyArgThrLeuGly 300  
 1138 TGGATCCCATGTAGTCCCTTTCAGTCTCGAGGCGCAACCTTGGGG 1187  
 301 LeuLeuValLysArgLeuGluLysGlyGlyLysAlaGluHisGluAsnLe 317  
 1188 TTTATGATAAAGATGGATGGAAAGTGTAAAGTGAACATGAAATTC 1237  
 317 uPheArgGluAsnAspCysLleValArgLleAsnAspGlyAspLeuArg 334  
 1238 TTTTCGTGAGATGATTCATGTCATGTCAGATTAATGATGGCAGACTTCGAA 1287  
 334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350  
 1288 ATAGAAGATTTGAACAAAGCAACATATGTTTCCCAACGATGGTACA 1337  
 351 ProLleLleTrpPheHisValAlaProAlaAlaAsnLysGluGlnTrpG1 367  
 1338 CCCATCATTTTGGTCCATGTGTTCTTGCAGCAAAATTAAGAGCAGTATGA 1387  
 367 uGlnLeuSerGlnSerGluLysAsnAsnYTrpTrpSerSerArgPheSer 384  
 1388 ACAACTATCCCAAGTGAAGAAACAATTAATCAAGCCGTTTATGCC 1437  
 384 roAspSerGlnTrpLleAspAsnArgSerValAsnSerAlaGlyLeuHis 400  
 1438 CTGACAGCCAGTATATGCAAAAGAGAGTGAACAGTGCAGGCTTCAC 1487

```
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
|||||
1488 ACGGTGAGAGAGAGACCCCGCATGMAACACCCGCTGACGAGATAGACTC 1537
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
|||||
1538 TCACCTCAGACTACCTCATAGCCGACACCCCTGGGAAAAACACCATCCG 1567
434 lArProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
|||||
1588 CTCACGCTGGGCACTCCAGATGTATTAGTAGACTGTAGAGAGTGGGT 1637
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
1638 TATATACACCAAAAAATAGGCAAGAGGCTTAATATCCAGCTTAAGAAGG 1687
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
|||||
1688 TACAGAAAGTTTGGATTCAGCATCACTCCAGAGATGTAAACAATACGTC 1737
484 lYSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
|||||
1738 GCTCAGCTCCAACTATGTGAAAAACATTCCTCCCGGGGGGGCCCATTT 1787
501 GluAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517
1788 CAGAGTGGCCGACTTAAGCGAGAGACAGACTTATAGAGGTAAATGGAGT 1837
517 lAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerT 534
|||||
1838 AGATTATGTGGCAAAATCCCAAGAGAAAGTGTTCCTGCTGTGAAAGCA 1887
534 hTrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
1888 CCAAGATGGAAGGAAGCTGTAGGCTTGTGCTTGTGCGCAAGGAAGAGCC 1937
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
1938 TTCACCCCAAGGGAAGTGAATGCGAGACCAGACGATGCAGATGCCAAA 1987
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584
|||||
1988 AGAAACGAAAGCAGAAAGATGAGATATGTCTTCAACCTGATGGCACCA 2037
584 rGluPheLeuThrPheGluValAlProLeuAsnAspSerGlySerAlaGly 600
2038 GGGATTTTCGACATTTGAAGTCCCACTTAATGATTCAGAGATCGCAGGC 2087
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
|||||
2088 CTTGGTGTCAGTGTCAAAAGGTAAACCGGTCAAAAGAAACCAACGCAGATT 2137
617 uGlyIlePheValLysSerIleIleAsnGlyValAlaLysSerLysAspG 634
|||||
2138 GGGAACTTTGTCAAGTCCATTATTAATGAGGAGGAGCAGCATCTAAAGATG 2187
634 lYArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysLeu 650
2188 GAAGGCTTCGGGTGAATGATCACTGATAGCAGTAATAATGGAGATCCCTG 2237
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
|||||
2238 TTGGGCAAGCAAAACCAAGATGCCATGGAAAACCTTAAGAGGTCTATGTC 2287
667 rThrGluGlyAsnLysArgLysMetIleGlnLeuIleValAlaArgArgI 684
|||||
2288 TACTGAAGCGCAATTAACGAGATGATCCAGCTTATTTGTGCAGAGAGAA 2337
684 lSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700
2338 TAAGCAAGTGAATGAGCTGAAGTCACTGGGAGCCGCCCTGGAGACTGAG 2387
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717

|||||
2388 CTGCCCAATTGAACGCGTTGTGATGATAGACAGCAAGAAATTTCCATTTC 2437
717 rLeuThrSerGlyIleGluGluLeuAspGluSerProSerArgAsnAla 734
2438 CCTCTACAGTGGGATTGAGGGCTTGTGATGAATCGCCGACGAAATGCTG 2487
734 lAlaLeuSerArgIleMetCylGlyLysSerGlyLysTyrGlnLeuSerProThr 750
|||||
2488 CCCACAGTAAAGATATG.....GGTAAATACAGCTTCCCTCCTACA 2528
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
2529 GTGAATATGCCCCCAAGATGACACTGTATATAGAAGATGACAGGTTGCC 2578
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784
2579 AGTCTTCTCTCCACATCTCTCTGACCACTCCTCTTCCAGCTCCCATATG 2628
784 sPValGlyPheValThrAlaAspAlaGlyThrTyrPalAlaLysAlaAlaIle 800
2629 ATGTGGGGTTTGTACGCGACATGCTGTACTTGGGCCAAGCTGCATTC 2678
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
2679 AGTGATTCAGCGCACTGCTTTGTAGTCCAGATGTTGATCCAGTTCCTTGC 2728
817 aPheGluArgGluGlyPheGlyArgArgLysSerMetSerGluLysArgThrL 834
2729 TTTTCACAGAGAAAGATTTGGACGTCAGAGTATGTCAGAAAAAGCACAA 2778
834 ySclnPheSerAspAlaSerGlnLeuAspPheValLysTyrArgLysSer 850
2779 AGCAATTTTCAGATGCCAGTCATTTGATTTGCTTAATAACAGAAAAATCA 2828
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
2829 AAAAGCATGGAATTAAGTATAGCTGACGACGACATCAATCAATACAGTCA 2878
867 pAspGluLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
2879 TGACCAAGAAACAGATTTCTCCACGACGAGATGTGGGTCTTCTCCCTGGGTC 2928
884 euLysLysSerSerSerLeuGluSerLeuGlnThrAlaValAlaGluVal 900
2929 TGAAGAAGTCAAGCTCATTTGAGAGTCTGCAGACCGCAGTTCGCGAGGTG 2978
901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleArg 917
2979 ACTTTGAATGGGGAATATCTTTCATCTGTCACGCGCGCGGATPAATCAG 3028
917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyrA 934
3029 AGGCAAGGGATGCAATGAGACTTCAGAGCTGCCATGCACAAATCTTATG 3078
934 sPLysProAlaValAlaAspAspAspGluGlyMetGluThrLeuGluGlu 950
3079 ATAAACCGCGGTAGATGATGATGAAGCATGAGAACCTTGGAAATAA 3128
951 AspThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlaSe 967
3129 GACACAGAAAGAAAGTTCAAGATCAAGGAGAGAGAGCTGTATCTCACAGCCAG 3178
967 rAspGluProSerHisSerLeuGluArgGluMetAsnGlyAsnGlnGlu 984
3179 TGATCAGGCTTCCCACTCTCTGAGAGACACAAATGAATGAACCAAGAGA 3228
984 ySclAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000
3229 AAGGTGAATGAAGCTGATGAATAAAGATTAACCTGGAAGAAAGAAAGAG 3278
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
|||||
```

```
3279 AAAGATAGATAGAGAGAGATATAAATGAAAGCCAGAGAGATCT 3328
1017 uTysgLyLeuGlyAspMetPheArgPheGlyHisArgLysAspSPL 1034
      |||||||
3329 GAAGGCGCTTGAGAGACATGTCACGCTTGCCAACTGAAAGCCGAGAGA 3378
      ::
1034 ys 1034
      ::
3379 GA 3380

seq_name: gb_pr:AF332592

seq_documentation_block:
LOCUS AF332592 2967 bp mRNA linear PRI 16-NOV-2001
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein
ACCESSION AF332592
VERSION AF332592.1 GI:14579308
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 2967)
  Fang, C.M. and Xu, Y.H.
  Down-regulated expression of atypical PKC-binding domain deleted
  Cell Res. 11 (3), 223-229 (2001)
JOURNAL MEDLINE PUBMED
21520080
1162408
REFERENCE
  2 (bases 1 to 2967)
  Fang, C. and Xu, Y.
  Exon/Intron Structure And Splicing Variants of A Novel Human
  Polarity Gene, hasip
  Unpublished
  3 (bases 1 to 2967)
  Fang, C. and Xu, Y.
  Direct Submission
  Submitted (29-DEC-2000) Laboratory of Molecular and Cellular
  Oncology, Institute of Biochemistry and Cell Biology, Shanghai
  Institutes for Biological Sciences, Chinese Academy of Sciences,
  350 Yue-Yang Road, Shanghai 200031, P. R. China
JOURNAL MEDLINE PUBMED
21520080
1162408
FEATURES
  source
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="10"
    /map="10p11.2"
    1..2967
    /note="ASIPSB: alternatively spliced, similar to GenBank
    accession Number AF196186"
    /codon_start=1
    /product="atypical PKC isotype-specific interacting
    protein short variant b"
    /protein_id="AAK69192.1"
    /db_xref="GI:14579309"
    /translation="MKTVCFGRRVVPCGDHMYVSLIQAVTRKATAKADPNY
    WQVHRLEHGDGILDDILCDVADKDIYAVEDPDHGGGTASSTGTSPSE
    IFSELCGTNNVASAPQYQATSEIEVTPSVLRAMPULVRRSSDPALIGISTVDSNF
    SSEPSRKNPTRRSTTAGTAKONTAGSPKTCRKDKDENRSIPRPTSMNSMNFORDNA
    RSSLSAHPVAVGKLEKQEDDEDEDEDSRYEVGHADTGEHFPRLKADDEDIVLT
    VPNDGCPGLIHVVFSAKGRTGLILYKLEKGAHEHLEHRENDICVIRINDGRLN
    RREBOAHMRQAMRPIIWFHYVPAANKEOYEOISQSRKNYSSRSPPSOYTDNR
    SVNSAGLHVORAPRLNHPPEQIDSHSRLPASHASQKPPASAPQNVSTVYSSG
    YNTKIKGRILQDKCTEGIGPSITSRDYTGGSAPLYKNIIPRGAIQDGRILKAG
    DRLIENGVDLVGKSGQEEVYSLRSTKMGVTSLLVFEQEAHFPRLEKADEDIVLT
    PGDTRELFEEVVDNSAGAGVSVKGRNSKENADIGIPKVISINSGAASKDGRIR
    VNOOLIAVNGESLGTKNODAMETLRKSKSTEGNRGAIQIIVARISIKCKELKSPS
    PREPELPIETALDDRERIRISHLSYSGIEGLDESPRNALRIKGTOLSPYVMPD
    DVIYIIDRLLPVLPPLSDOSSSHSDVGEFTYADAGWAAALISDASDGLSPVMPD
    VLAFOEGFGROIADETKLINTVDQKASPSDPSGILKSSLSLQTAVALVPL
    MDIPIHRRPRIIRGRCNESFRAIDKSYDKPAVDDDKMETLEDEEESRGR
    ESYSTASDQPSHSLEROMNGNEKDKTRKKDKTGAEKKRDKRDKKKAKKGMK
```

```
BASE COUNT      GUDMFSLAKLPKRR"
ORIGIN          890 a      653 c      776 g      648 t

alignment_scores:
  Quality: 5014.00      Length: 1034
  Ratio: 5.096
  Percent Similarity: 95.164      Percent Identity: 94.681

alignment_block:
US-09-757-781-2 x AF332592 ..
Align seq 1/1 to: AF332592 from: 1 to: 2967

1 MetLysValThrValAlcysPheGlyArgThrArgValAlaValProCysGI 17
  |||||||
1 ATGAAAGTGACCGTGTGCTGACGACGACCCGGGTGCTGCTGCGG 50

17 YAspGlyHisMetLysValPheSerLeuIleGlnAlaValThrArgT 34
  |||||||
51 GCAGCGCCACATGAAGAGTTTTCAGCCTCATCCAGCAGCGGTGACCGCT 100

34 YArgLysAlaIleAlaLysAspProAsnTYTTPILeglnValHisArg 50
  |||||||
101 ACCGGAAAGCCATCGCCAAAGATCCAAACTGATGATACAGTCAATCGC 150

51 LeuGlnHisGlyAspGlyLYTleLeuAspLeuAspIleLeuCysas 67
  |||||||
151 TTGGAACATGGAGATGGAGAGAACTAGACCTTGATACATCTTGTGTA 200

67 PValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84
  |||||||
201 TGTACGACAGCATAAAGACAGACTGTAGCAGTGTGATGAGACAGACATC 250

84 rOHLSHLSGLYLYAspGlyTYHSerAlaSerSerThrGlyThrGlnSer 100
  |||||||
251 CACATACAGGAGGTGATGCGACACAGTCCAGTTCACGGGTGCCAGAGC 300

101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG 117
  |||||||
301 CCAAGATATTGTGTAAGTGTGACACCAACAAATGCTCAGCCTTCA 350

117 nProTyGlnAlaThrSerGluIleGluValThrProSerValIleuArg 134
  |||||||
351 GCCTTACCAAGCAACAAGTAATTAGGTCAACCTTCAGTCCCTTCAG 400

134 LAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
  |||||||
401 CAATATGCTCTTCATGTTCAGACGACAGTAGACCCAGCTCAATTGCG 450

151 LeuSerThrSerValSerSerAspSerAspPheSerSerGluGluProSer 167
  |||||||
451 CTCTCAGCTTGTGTCAGTATGATTAATTTTCTCTGTAAGGCTTCAAG 500

167 GlyAsnProThrArgTTPSerThrAlaGlyPheLeuLysGlnAsnT 184
  |||||||
501 GAAAAATCCACACAGCTGTCACAAACAGCTGGCTTCTCAGCAGACACA 550

184 hPAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTYArg 200
  |||||||
551 CTGCTGGGAGTCTTAAACCTCGCAGCAGGAAAGTAAGTAACACTACAGA 600

201 SerLeuProAlaGAspPThrSerAsnTTPSerAsnGlnPheGlnArgAsp 217
  |||||||
601 AGCTCCCGCGGAGTACTAGTAACCTGCTAACCACTTTCAGAGACACA 650

217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTTPLeuG 234
  |||||||
651 TGTCTGCTGCTCTGTGATCCAGTACCCAAAGTGGGCAAGTGGCTG 700

234 LuYsgLingLugLInAspGlyAspGlyTYHrGluGluAspAsnSerArgVal 250
  |||||||
```



```

701 AGAAACAAGACAGATGAGATGGACACAGAGAGATTAACACTCGTGT 750
251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPhse 267
751 GAACCGTTGGACATGCTGACAGGGTTTGGACATTAACCAACTTTC 800
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLeuProL 284
801 TCTGGATGATGTGTAAAGCTCTAGAACTCCCAACGATGAGAGGCTC 850
284 euGlyIleHisValIleProPheserAlaArgGlyArgThrLeuGly 300
851 TGGGAATCCATGTAGTGGCTTTTCAGTCTGAGCGGCGACAGACCTGGG 900
301 LeuLeuValLysArgLeuGlyLysGlyLysAlaGluHisGluAsnLe 317
901 TTATTGTATAAAGATTTGGAAAGGTGGTAAGCTGAACATGAAATATCT 950
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
951 TTTTCGTGAAATGATGTGATGTCAGATTAATGATGGCGACCTTCGAA 1000
334 snrArgArgPheGluGluAlaGluHisMetPheArgGluAlaMetArgThr 350
1001 ATGAAAGATTTGACACAGCAACATATGTTTGGCCAAAGCCATGCGTACA 1050
351 ProIleIleTrpPheHisValIleProAlaAlaAsnLysGluGlnTrpG 367
1051 CCCATCATTTGGTTCCATGTGGTTCCGCGCAATTAAGACAGATATGA 1100
367 uGlnLeuSerGlnSerGluLysAsnAsnTrpTrpSerSerArgPheSerP 384
1101 ACAACATATCCCAAGAGTGAGAAACAATTAATCTCAAGCCGTTTATGCC 1150
384 roAspSerGlnTrpIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
1151 CTGACAGCCGATATATGACACAGAGTGTAACAGTGCAGGGCGTTTCAC 1200
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
1201 ACGGTGACAGAGACACCCGCACTGAACCCAGCCCTGAGCAGATATGACTC 1250
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
1251 TCACCTAAGACTACCTCATGACCCACACCCCTCGGAAACCAACATCCG 1300
434 laProAlaSerAlaProGlnAsnValPheSerThrValSerSerGly 450
1301 CTCACAGCTCGGCACTCAGATGATTTAGTACGACTGTAAACAGTGGT 1350
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
1351 TATAACACCAAAAAAATAGCAGAGGCTTAATATCCAGCTTAAGAAAGG 1400
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484
1401 TACAGAGAGGTTTGGATTCAGATCACTTCCAGAGATGTAAACAATAGTG 1450
484 lysSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
1451 GCTCAGCTCCAAATCTATGTGAAAAACATTTCCCGCGGGGGCGCCATT 1500
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVal 517
1501 CAGGATGGCGCACTTAAGCAGAGACAGACTTAAGAGGTAATGAGT 1550
517 lasPheValGlyLysSerGlnGluGluValValSerLeuLeuArgSerT 534
1551 AGATTTAGTGGCAAAATCCCAAGAGGAAATGTTTCTGCTTGAGAGACA 1600
534 hrLysMetGluGlyThrValSerLeuLeuValPheArgGlnLysAspAla 550
1601 CCAAGATGGAGAACTGTGAGCTTCTGCTTTCGCCAGAAAGACGCC 1650
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
1651 TTCACCCCAAGGAACTG..... 1668
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584
1669 .....AAAGCAGAAAGATGAGATATTGTTCTTACACCTGATGGCACCA 1711
584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
1712 GGGAAATTTCTGACATTTGAAAGTCCCACTTAATGATTCAGAGATGCGAGC 1761
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
1762 CTGGTGTGAGTGTGCACAAAGGTAAACGGTCAAAAAGAAACCAAGCATTT 1811
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
1812 GCGAATCTTTGTCAAGTCCATTTATTAATGAGAGCAGCATCTAAAGATG 1861
634 lyArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650
1862 GAAGGCTTCGGGTGAATGATCACTGATAGCAGTAAATGGAGAAATCCCTG 1911
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
1912 TTGGGCAAGAACAAACCAAGATGCCATGGAAACCTTAAGAAAGTGTATGTC 1961
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684
1962 TACGAAAGCCAAATTAACGAGAAATGATCCAGCTTATTTGTTCAAGAGAA 2011
684 leSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700
2012 TAAGCAAGTGCATGAGCTGAAGTCACTGAGCGACCCCTGGACCTGAG 2061
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
2062 CTGGCCATTGAACACAGGTTGGATGATAGAGAACGAAGAAATTTCCATTC 2111
717 rLeuTrpSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAlaA 734
2112 CCTTACAGTGGGCAATGAGGGCTTGATGAATCCCGCAGCAGAAATGCTG 2161
734 laLeuSerArgIleMetGlyGluSerGlyLysTrpGlnLeuSerProThr 750
2162 CCTCAGTAGGATTAATG.....GCTAAATACCAAGCTGTCCCTTACA 2202
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
2203 GTGAATATGGCCCAAGATGACACTGTCATTAATTAAGATGACAGGTTGCC 2252
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784
2253 AGTGCCTTCTCCACATCTCTGTGACAGTCTTCCAGGCTCCCATGATG 2302
784 spValGlyPheValThrAlaAspAlaGlyThrThrPalalysAlaAlaIle 800
2303 ATGTGGGGTTTGTGACGCGCAGATGCTGTACTGTGGCCAAAGCTGCATTC 2352
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
2353 AGTATTCACCGCACTGCTCTTGAAGTCCAGATTTGATCAAGTTCCTGC 2402
817 apPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
2403 TTTTCAACGAGAAAGATTTTGACCTCAG..... 2430
834 ysgInPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
2430 ..... 2430

```

```

851 lysSerMetaspLeuGlyIleAlaAspGluThrIleuAsnThrValAs 867
2431 .....ATAGCTACGACGACTAAACTCATACAGTGA 2462
867 PASpGlnIysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
2463 TGACCAGAAAGCAGGTTCTCCAGCAGAGATGGTGGTCCCTCCGCTC 2512
884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGlu 900
2513 TGAAGAACTCAAGCTCATGGAGAGTGGCAGACCGCAGTTGCCGAGCTG 2562
901 ThrIleuAsnGlyAspIleProPheHisArgProArgProArgIleIle 917
2563 ACTTGAAATGGGAGATTCCTTCATCGTCACGGCCGGGAGAAATCAG 2612
917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerIyr 934
2613 AGCGAGGGAGATGCAATGAGAGCTTCAGAGCTGCCATCGACAAATCTTATG 2662
934 sPlySProAlaValAspAspAspGluGlyMetGluThrLeuGluGlu 950
2663 ATAAACCCGGCGTAGATGATGATGATGATGATGATGATGATGATGATGAT 2712
951 AspThrGluGlnSerSerArgSerGlyArgGlnSerValSerThrAla 967
2713 GACACAGAAAGAAAGTCAAGATCAGGAGAGAGTCTGTATCCACAGCAG 2762
967 rAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlyAsnGlnGlu 984
2763 TGATCACCCTCCCACTCTCTGGAGAGACAAATGANTGAAACCAAGAGA 2812
984 yGlyGlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLys 1000
2813 AAGGTGTAAGACTGATAGAAAAAGATTAACCTGAAAAAGAAAGAG 2862
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMet 1017
2863 AAGATATAGATATAGAGAAAGATTAATTAAGCCAGAGAGGAAATGCT 2912
1017 uLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAsp 1034
2913 GAAGGCTTGGGAGACATGTTTCAGCTTCGCAAACTGAAGCCGAGAGAA 2962
1034 ys 1034
2963 GA 2964
seq_name: gb_pat:BD004905
seq_documentation_block:
LOCUS BD004905 2718 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel genes encoding protein kinase or protein phosphatase.
ACCESSION BD004905
VERSION BD004905.1 GI:18632866
KEYWORDS JP 03074935-T/2.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2718)
Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Makamatsu,A., Nagai,K., Otsuki,T.,
Funahashi,S., Seno,C. and Nezu,J.
Novel genes encoding protein kinase or protein phosphatase
Patent: JP 03074935-T 2 30-JAN-2001;
HEIIX RESEARCH INSTITUTE, TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA,
KOJI HAYASHI, KAORU SAITO, JUNICHI YAMAMOTO, SHIZUKO ISHII, OMOYASU
SUGIYAMA, AI WAKAMATSU, KEIICHI NAGAI, TETSUJI OTSUKI, SHINICHI
FUNAHASHI, HIROKI SENO, JUNICHI NEZU
OS Homo sapiens (human)
PN JP 03074935-T/2
PD 30-JAN-2001

```

```

PF 28-JUL-2000 JP 2000005061
PR 29-JUL-1999 JP 99P 248036,11-JAN-2000 JP 00P 118776 PR
02-MAY-2000 JP 00P 183767,09-JUN-2000 JP 00P 241899 PR
18-OCT-1999 US 60/159590,17-FEB-2000 US 60/183322 PI TOSHIO
OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI, PI KAORU SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI,TETSUJI OTSUKI,SHINICHI FUNAHASHI,CHIHIKI SENO,
PI JUNICHI NEZU
PC C12N15/12,C12N9/12,C12N9/16,C12N1/15,C12N1/19,C12N1/21, PC
C12N5/10,C07K16/40,
PC C12Q1/68,C12Q1/42,C12Q1/48
CC
FH
FT CDS Location/Qualifiers
1..2718
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 840 a 586 c 688 g 604 t
ORIGIN
FEATURES
source
1..2718
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
alignment_scores:
Quality: 4409.00 Length: 909
Ratio: 5.068 Gaps: 2
Percent Similarity: 95.710 Percent Identity: 94.829
alignment_block:
US-09-757-781-2 x BD004905 ..
Align seg 1/1 to: BD004905 from: 1 to: 2718
126 GluValThrProSerValLeuArgLysAsnMetProLeuHisValArgAr 142
3 GAGGTCAACCTTCAGTCCCTTCGAGCAAAATATGCTCTTCATGTTCCGCG 52
142 gSerSerAspProAlaLeuIleGlyLeuSerThrSerValSerAspSera 159
53 CAGTAGTACCCAGCTTAATTTGGCTCTCCACTTCTGTCTGATATAGTA 102
159 snPheSerSerGluGluProSerArgLysAsnProThrArgTrpSerThr 175
103 ATTTTTCCTCGAAGAGCCCTTCAGGAAAAATCCACACGCTGGTCAACA 152
176 ThrAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysThrCysAs 192
153 ACAGCTGCTTCCTCCAGCAGACATGCTGAGAGTCCCTAAAGCTCGCA 202
192 PArgLysLysAspGluAsnTrpArgSerLeuProArgAspThrSerAsn 209
203 CAGGAAAGAAAGATGAAACTACAGAAAGCTCCCGCGGAGATGACTAGTACT 252
209 rPserAsnGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSer 225
225 GGTCTAACCAATTTTCAGAGACATGCTCGCTCGCTCGAGTCCAGT 302
225 GGTCTAACCAATTTTCAGAGACATGCTCGCTCGCTCGAGTCCAGT 302
226 HisProMetValGlyLysTrpLeuGlnLysGlnGlnGlnAspGluAsp 242
303 CACCAATGTTGGGCAAGTGGCAGAGAGAAACAGAACAGATGAGATGG 352
242 yThrGluGluAspAsnSerArgValGluProValGlyHisAlaAspThrG 259
353 GACGAAAGAGATTAACAGTGTGTTGAACTGTTGGACATGTCGACACGG 402
259 LysLeuGlnHisIleProAsnPheSerLeuAspAspMetValLysLeuVal 275
403 GTTTGGAGCATATACCACTTTTCTGTGATATATGTAAGCTCGTA 452
276 GluValProAsnAspGlyGlyProLeuGlyLysIleValValProPheS 292
453 GAAGTCCCAAGCATGAGAGGCTCTGGGAATTCATGATGCTTTCAG 502
292 rAlaArgGlyGlyArgThrLeuGlnLeuValLysArgLeuGluLysG 309

```

```

|||||
503 TGTTCAGAGCGGCGAACCCTGGGGTTATTAGTAAAAAGATTGGAGAAAG 552
309 IYGLIYSLAAGIUNHISGLUASNLEUPHEARGIUNASNAPCYSLIIVAL 325
553 GTGGTAAGCTGAACATGAAAATCTTTTGTGAGAAATGATTCATGCTGC 602
326 ARGILEASNAAPGLYASPLEUARGASNAARGARGPHEGLIUNALAGINH 342
603 AGGATTAATCATGTGCGCCTTCCAAATAGAAAGATTGAAACAGCACAA 652
342 SMETPHEARGSLALAMETARGTHRPROILEIETRPHEHISVALYALP 359
653 TATGTTGCGAAGCCATGCGTACACCATCATTTGGTTCATGTGGTTC 702
359 ROALAAASNLYSGIUNTYRGLIUNLEUSERGINSEARGIULYSASN 375
703 CTCACAGAAATAAGAGCACTATGAAACAATCCCAAAAGTGAAAGAAC 752
376 ASNTYTYRSESRARGPHESERPPOASPSEGLINTYRILEASPNAR 392
753 AATTACTATTCAAGCCGTTTACGCCGACAGCCAGTATTTGACACAG 802
392 GSERVALASERIALGLYLEUHNSTHRVALGINARGALAPROARGLEUA 409
803 GAGTGTGAACAGTGCAGGGCTTCACAGGTGCAGAGAGCACCCGACTGA 852
409 SNHSPROFPROGLIUNILEASPSERHISSEARGLEUPROHISSEARA 425
853 ACACACCGCGTGAGCAGATGACTCTCACAGACACTACTCATAGCGCA 902
426 HISPROSEGLYLYSPROPSERAPALAPROALASERIALPROGLIUNA 442
903 CACCCCTCGGGAAAAACCACTCCGCTCCAGCCTCGGCACCTCAGATGT 952
442 IPHSESRTHTHRVALSERSEGLYTYRASNTHRYLYSLILEGLIYLSA 459
953 AATTAGTACACGTGTAAGCAGGTGTATTAACCAAAAAAATAGGCAAGA 1002
459 RYLEUASNILEGINLEUYSLYSGLYTHRGLIUNGLYPHESERILE 475
1003 GGGTTAATATCCAGCTTAAGAAAGTACAGAAAGTTGGGATTCAGCATC 1052
476 THRSEARGAPVALTHRILEGILYSEARALPROILETYRVALYLSAS 492
1053 ACTTCCAGAGATGAACAATAGGTGCTCAGCTCCAAATCTATGTGAAAA 1102
492 NILELEUPROARGLYALALALILEGINASPGLYARGLEULYSALGLYA 509
1103 CATTCCTCCCGGGGGGGCGGCATTCGAGATGGCCGACTTAAGGCAAGAG 1152
509 SPARGLEULILEGLIUALASNGLYVALASPLEUVALGLIYSSERGLINU 525
1153 ACAGACTTATAGAGTAAATGAGATGATTTAGTGGCAAAATCCCAAGAG 1202
526 GLUVALVALSERLEUENARGSESRTHLYSMETGLIUNLYTHRVALSER 542
1203 GAAGTTGTTTCGCTGTGAGAACGACCAAGATGGAAGAACTGTGAGACT 1252
542 ULEUVALPHEARGINGIUNASPLAPHEHISPROARGIUNLEUASNLAG 559
1253 TCTGGTCTTTCGCCAGAGAGCGCTTCCACCCAAAGGGAACGTGAATCAG 1302
559 IUPROSEGINMETGLINILEPROLYSGIUNTHRYLYSALAGIULASPLASP 575
1303 AGCCAGACCAAGATGACATTCCAAAAAGAAAGCAAGCAAGAGAGAGAT 1352
576 ILEVALLEUTHRPROASPGLYTHYRGLUPHEUETHRPHGLIUALPR 592
1353 ATGCTTCTACCTGATGGCAGCAGGAATTTCTGACATTTGAAGTCC 1402
592 OLEUASNAPSESGLYSERIALGLYLEUGLYVALSERVALYSGLYASNA 609
|||||
1403 ACTTAGTATTCAGGATTCGACGGCTTGGTGTCAAGTGTCAAAAGTAAAC 1452
609 RYSERLYSLIUNASNHISALASPLEUGLYILEPHEVALYSSERILE 625
1453 GGTCAAAAGAGAAACACAGCATTTGGGAATCTTTGTCAACATCATATAT 1502
626 ASNGIYGLYALALASERLYSASPGLYARGLEUARGVALASNAPSEINLE 642
1503 AARGAGAGCAGCATCTTAAAGATGAAAGGCTTGGGTGAATGATWCACT 1552
642 UILEALVALASNGIYGLIUNSERLEUENGLYLYSTHRASNGLINAPLAM 659
1553 GATAGCAGTAAATGAGAAATCCCTGTTGGGCAAGACAAACCAATGCCA 1602
659 ETGLIUNTHREUARGARGSEMETSESRTHRGILYASNLYSARGLYMET 675
1603 TCGAAACCTTAAGAAAGSTCTATGTCTACTGAAGCAATTAACGAGAAATG 1652
676 ILEGINLEULEVALALAAARGARGILESERLYSCYASNGIULEULYSE 692
1653 ATCCAGCTTATTTGTGCAAGAGAAATAGCAAGTCAATGAGCTGAAGTC 1702
692 RPROGLYSEPROPROGLYPROGLIUNLEUPROILEGLIUNTHRALASNA 709
1703 ACTTGGAGCCCGCCCTGAGCTTGAGCTGCCATTGAAACAGCCTTGATG 1752
709 SPARGIUNARGARGILESERHISSERLYSEARGIYILEGLIUNLEU 725
1753 ATAGAGAAACGAAGAATTTCCCATTCCTCTACAGTGGGATTGAGGGGCTT 1802
726 ASPLIUSERPROSERARGASNALALALEUSERARGILEMETGLYGLUSE 742
1803 GATGAATCGCCACAGAAATGCTGCCCTCAGTAGAGATAATG..... 1844
742 RGLIYLYSTYRGLIUNLEUSERPROTHRALASMETPROGLINASPSPTHY 759
1845 .GGTAAATPACAGCTGTCCCTACAGTGAATATATCCCAAGATGACACTG 1893
759 ALILEILEGLIUNASPSARGLEUPROVALLEUPROPROHISLEUSERASP 775
1894 TCATTATAGAAAGATGACAGGTGGCAGTCCCTCCACATCTCTGTGAC 1943
776 GLINSESRSESRSESRHISASPSAPVALGLYPHEVALTHRALASPAL 792
1944 CAGTCTCTTCCAGCTCCATGATGATGTGGGGTGTGAGAGGAGATGC 1993
792 AGLYTHRTPRALYSLALALALILESEASPSERIALASPCYSSERLEUS 809
1994 TGGTACTTGGCCAAAGGCTGCAATCAGTATTCAGCCGACTGTCTTTGA 2043
809 ERPROASPVALASPPROVALLEUALAPHEGINARGIUNGLYPHEGLYARG 825
2044 GTCCAGATGTTGATCCAGTCTTCTGCTTTCAACGAAAGAAATTTGGAGCT 2093
826 GLINSEMETSESGIULYSARGTHRYLYSGLINPHESEASPALASERGLINLE 842
2094 CAGACT..... 2099
842 UASPPHEVALYLYSTHRARGLYSERLYSEMETASPLEUGLYILEALAA 859
2100 .....G 2100
859 SPGLIUNTHRYLYSLEUASNTHRYVALASPSAGIUNLYSALAGLYSERPROSE 875
2101 ACAGAGACTAAATCAATPACATGATGATGACAGAAAGCAAGTTCGCCAGC 2150
876 ARGASPVALGLYPROSERLYSEUGLYLEULYSLYSSERSESRLEUGLUSE 892
2151 AGAAGATGGGTCTTCCCTGCTGCTGAAAGAAAGTCAATCTATTTGGAAG 2200
892 RLEUGIUNTHRALVALALAGIUALALTHYLEUASNGLYASPLIETPROPHEN 909
2201 TCTGCAGACCCGACTTGGCGAGGTGACTTTGAATGGGATTTCTCTTCC 2250

```

909 isatgproatrgproatrgilleatrgglyargglycysancluserphe 925  
|||||  
2251 ATCTGTCACGGCGCGGATCAATCAGAGGCGATGCAATGAGAGCTTC 2300  
926 ATGATAAaalleasplyserTyrAsplysProAlaValAspAspAsp 942  
|||||  
2301 AAGGTGCCATCGACAAATCTTATGATTAACCCGGGATGATGATGATCA 2350  
942 pgluglYmetGluThrLeuGluGluAspThrGluGluSerSerArgSerg 959  
|||||  
2351 TGAAGCATGAGGACCTTGGAAGAACACAGAAAGTTCACAGATCAG 2400  
959 lYargGluSerValSerThrAlaSerAspGlnProSerHisSerLeuGlu 975  
|||||  
2401 GGAGAGAGCTGTATCCACAGCCAGTATGATCAGCTTCCACTCTCGAG 2450  
976 ArgGlnMetAsnGlyAsnGlnGluGlyAspGlyThrAspArgGlySyl 992  
|||||  
2451 AGACAAATGAATGGAACCAAGAAAGGATGATTAAGACTGATGAAAAA 2500  
992 sAspLysThrGlyGlyGluGlySylAspArgAspLysGlyLysAspL 1009  
|||||  
2501 GGATTAACACTGGAAAAAGAAAGAAAGATAGACATTAAGAGAAAGATA 2550  
1009 ysmetLysAlaLysLysGlyMetLeuGlyLeuGlyAspMetPheArg 1025  
|||||  
2551 AAATGAAAGCCAAAGAAAGGAATGCTGAAGGCGCTGGAGACATGTTCA 2600  
1026 PheGlyLysHisArgLysAspAspLys 1034  
|||||  
2601 CTTGCCAAACTGAAGCCCGAGAGAGA 2627  
seq\_name: gb\_pr:AK027735  
seq\_documentation\_block:  
LOCUS AK027735 2718 bp mRNA linear PRI 15-MAY-2001  
DEFINITION Homo sapiens cDNA FLJ14829 f1s, clone OVARC100945, moderately  
similar to Rattus norvegicus mRNA for atypical PKC specific binding  
protein.  
ACCESSION AK027735  
VERSION AK027735.1 GI:14042633  
KEYWORDS cDNA, capping, f1s (full insert sequence).  
SOURCE Homo sapiens ovary, tumor tissue cDNA to mRNA, clone\_l1b:OVARC1  
clone:OVARC100945.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,  
Wagatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,  
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,  
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,  
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,  
Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.  
TITLE NEDO human cDNA sequencing project  
REFERENCE 2 (bases 1 to 2718)  
AUTHORS Isogai,T. and Otsuki,T.  
JOURNAL Direct Submission  
SUBMITTED (10-MAY-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@h1.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing;  
Research Association for Biotechnology; cDNA library construction,  
5'- & 3'-end one pass sequencing and clone selection; Helix  
Research Institute (supported by Japan Key Technology Center etc.)  
and Department of Virology, Institute of Medical Science,  
University of Tokyo.  
FEATURES  
Source Location/Qualifiers  
1..2718

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="OVARC100945"  
/tissue\_type="ovary", tumor tissue"  
/clone\_l1b="OVARC1"  
/note="cloning vector: pME18SFL3"  
33..2630  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB55330.1"  
/db\_xref="GI:14042634"  
/translation="MPILHVRSSDPALIGISTSVSDNSESESESRKNPTRWTTAGF  
LKQNTAGSPKACDKKDENYRSLPRDTSNNSNOFORNASSSUSASHPMWKQKEOE  
ODEGCTEDNSRVEPVGHADTGLEHLPNFSLDIMVKLVLPNDGPGIIVVPSARG  
GRTGLGLVKRLEKGAHEHNFENDICIVRIDNDLRNRFEDQAHMFRQAMPTPII  
MEHVPAKNEOYEOLDSOSERNYNSRSFSPDSQYIDNRVNSGLTVORAPLNP  
PROIDSHRLPHSAHPSPKSPASAPASAPASAPASAPASAPASAPASAPASAP  
GLGFSSTRVYITIGSAPITYKNILPFGCAIIOGRKLAKGRLLIEVNGVDLVGSQEBY  
VSLRSTKMEGTVALVFROEDAHPRELNAEBSQOIPETAEDEDIVLTPDGTRE  
FLTFEVPILSDSGSAGLGVSKNSKRNHADLGIPIVSIINGKAASKDGRILRVNDLI  
AVNGESLIGRTNODAMETTLRRSMSTEGNKGMIOLIVARRISKNEELKSPGPGPL  
PIETALDDRERRTISHLSYSGIEGLDESPRNALSLRMGYOLSPYMPDOPVITE  
DDRLPVLPPLISDOSSSSSHDDVGFVTAADGTAKAKAISDADCSLSPDVPYAPOR  
EGRGRQIDERKLTIVDQKAGSPSRDYPGSLGIKKSLSLSLOTAAVEVTLNGLDIPR  
RPRRIIRGRGNCNFSFRAALDKSTDKPAVDDDCMGTSLLEDETESSRSRESVTS  
DQPSHSLERMQNGNQEKGDKTDRKKDTGKEKKRDKERDKMKAKKGMGLGDMPS  
LAKLKPEKR"  
BASE COUNT 840 a 586 c 688 g 604 t  
ORIGIN  
alignment\_scores:  
Quality: 4409.00 Length: 909  
Ratio: 5.068 Gaps: 2  
Percent Similarity: 95.710 Percent Identity: 94.829  
alignment\_block:  
US-09-757-781-2 x AK027735 ..  
Align seg 1/1 to: AK027735 from: 1 to: 2718  
126 GUAUATThrProSerValLeuArgAlaAsnMetProLeuHisValArgArg 142  
|||||  
3 GAGGTGCACACCTTCAGTCCCTCGAGCAAAATATGCCCTTCACATGTTGCGAGC 52  
142 gserSerAspProAlaLeuIleGlyLeuSerThrSerValSerAspSerA 159  
|||||  
53 CAGTAGTGACCCAGCTCTAATGGCTCTCCACTTCTGCAATGATAGTA 102  
159 snPheSerSerGluGluProSerArgLysAsnProThrArgTTPSerThr 175  
|||||  
103 ATTTTCTCTGTAAGAGCTTCAAGAAAAATCCACAGCGTGGTCAACA 152  
176 ThrAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysThrCysAs 192  
|||||  
153 ACAGCTGGCTTCTCCTCAAGCAGAACACTGCTGGAGTCCCTAAAGCCTGCGA 202  
192 parGlyLysAspGluAsnTyrArgSerLeuProArgAspThrSerAsnT 209  
|||||  
203 CAGGAGAAAGATGAATAACTACAGAGAGCTCCCGGGGAGTACTAGTAAC 252  
209 rPserAsnGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSer 225  
253 GGTATACCAATTTCAGAGAGACATGCTCGCTCTGTAGTGCACAT 302  
226 HisPrometValGlyLysTrpLeuGluGlyGlnGluGlnAspGluAspG 242  
|||||  
303 CACCCAAATGTGTGGCAGAGTGCAGAGAAACAAAGAACAGATGAGAGATGG 352  
242 yThrGluGluAspAsnSerArgValGluProValGlyHisAlaAspThrG 259  
|||||  
353 GACAGAAAGAGATTAACAGTCTGTGAACCTGTGTGACACTGTGACACGG 402

259 IyLeuGIuHISiIeProAsnPheSerLeuAspMetValIyLysLeuVal 275  
 |||  
 403 GTTGGAGCATATACCACTTTCTCTGATGATATGTAAGAGCTTCGTA 452  
 276 GluValProAsnAspGlyGlyProLeuGlyIleHisValValProPheSe 292  
 |||  
 453 GAAGTCCCAACGATGAGGAGGCTCTGGGAAATCAGTAGTGCTTTTCAG 502  
 292 rAlaArGgLyGlyArGThrLeuGIyLeuLeuValIySArGLeuGIyLysG 309  
 |||  
 503 TGTCTGAGGCGGAGAACCTTGGGGTATTAGTAAACATTGAGAGAAAG 552  
 309 IyGlyLySAlaGlyuHISgluAsnLeuPheArGgLyLysAsnAspCysIleVal 325  
 |||  
 553 GTGGTAAACTGMAACATGAAAATCTTTTCGTGAGATGATTCATTTGTC 602  
 326 ArgIleAsnAspGlyAspLeuArgAsnArgArgPheGluGlnAlaGlnH 342  
 |||  
 603 AGGATTAAATGATGGCACTTCGAAATAGAAAGATTGAAACCAAGCAACA 652  
 342 sMetPheArGgGlnAlaMetArGThrProIleIleTrpPheHisValValP 359  
 |||  
 653 TATGTTCCGCAAGCCATGGGTACACCATTCATTGGTTCCATGTGTTTC 702  
 359 rAlaAlaAsnLySgLyGlnIyTrGlyGlnLeuSerGlnSerGlyLysAsn 375  
 |||  
 703 CTGCACCAATAAAGAGAGATATGAACAACTATCCAAAGTGAGAAAGAC 752  
 376 AsnIyTrSerSerArGpPheSerProAspSerGlnIyTrIleAspAsnAr 392  
 |||  
 753 AATTACTATTCAGCCGTTTTACCTTGACAGCCAGCATATTTACAAACAG 802  
 392 gSerValAsnSerAlaGlyLeuHISThrValGlnArGAlaProArGLeuA 409  
 |||  
 803 GAGTGTGAACAGTGCAGGGCTTCACAGGGTGAGAGAGACCCGACTGA 852  
 409 snHisProGlyGlnIleAspSerHisSerArgLeuProHisSerAla 425  
 |||  
 853 ACCACCCGCTGAGCAGATAGACTCTCACCACCAAGACTACTCTCATAGCGCA 902  
 426 HisProSerGlyLysProProSerAlaProAlaSerAlaProGlnAsnVa 442  
 |||  
 903 CACCCCTCGGAAACACCAACATCGCTCCAGCTCGGACCTCAGAAATGT 952  
 442 lPheSerThrThrValSerSerGlyTyTrAsnThrLySlyIleGlyLysA 459  
 |||  
 953 ATTTAGTAGACTGTAGACAGTGGTTATACACCAAAAAATAGGCAGAGA 1002  
 459 rGlyLeuAsnIleGlnLeuLySlyGlyThrGlyGlnLeuGlyPheSerIle 475  
 |||  
 1003 GGCTTATATCCAGCTTAAGAAAAGTACAGAAAGTTTGGGATTCAGCATC 1052  
 476 ThrSerArGAspValThrIleGlyGlySerAlaProIleTyValIySAs 492  
 |||  
 1053 ACTTCCAGAGATGTAAACAAATAGGTGGCTCAGCTCCAAATCTATGTGAAAAA 1102  
 492 nIleLeuProArGgLyAlaAlaIleGlnAspGlyArGLeuLySAlaGlyA 509  
 |||  
 1103 CATTCTCCCGGGGGGGCGGCATTCAGATGGCCGACCTTAAGCAGAGAG 1152  
 509 sPArGLeuIleGlyuValAsnGlyValAspLeuValGlyLySserGlnGlu 525  
 |||  
 1153 ACAGACTTAAAGAGTAAATGAGATTTAGTGGCCAAATCCCAAGAG 1202  
 526 GluValValSerLeuLeuArgSerThrLyMetGlyGlyThrValSerIle 542  
 |||  
 1203 GAAGTGTTCCTGCTGTGAGAACACCAAGATGGAAGGAACCTGTGACCT 1252  
 542 uLeuValPheArGgGlnGluAspAlaPheHisProArGgLyLysAsnAlaG 559  
 |||  
 1253 TCTGCTCTTCCGCAAGAGACGCTTCACCAAGGAAGTGAATGACAG 1302  
 559 IuProSerGlnMetGlnIleProLySgLyuThrLySAlaGlyuAspGlyuAsp 575

|||  
 1303 AGCCAAGCCAGATGACGATTCCAAAGAAAGCAAGACGAGATGAGCAT 1352  
 576 lIleValLeuThrProAspGlyThrArGgLyPheLeuThrPheGlyuValP 592  
 |||  
 1353 ATTTGTTCTTACACCTGATGAGCACCAGGAAATTTCTGACATTTGAAGTCCC 1402  
 592 oLeuAsnAspSerGlySerAlaGlyLeuGlyValSerValLySgLyAsnA 609  
 |||  
 1403 ACTTAGTATTCAGAGATCTGACAGCCCTGGTGTGACAGTGCAAAGGTAAAC 1452  
 609 rGSerLySgLyAsnHisAlaAspLeuGlyIlePheValLySserIleIle 625  
 |||  
 1453 GGTCAAAAGAGAAACCAAGCAGATTGGGAATCTTTGCAAGTCCATTAAT 1502  
 626 AsnGlyGlyValAlaSerLySAspGlyArGLeuArgValAsnAspGlnLe 642  
 |||  
 1503 AATGGAGGAGACGACATCTTAAGATGGAAGGCTTCGGGTGATGATCAACT 1552  
 642 uIleAlaValAsnGlyGlyuSerLeuLeuGlyLySThrAsnGlnAspAlaM 659  
 |||  
 1553 GATAGCAGTAAATGAGAAATCCCTGTTGGGCAAGACAAACCAAGATGCCA 1602  
 659 eGlyuThrLeuArGArGSerMetSerThrGlyLysAluLySArGgLyMet 675  
 |||  
 1603 TGGAAACCCCTAAGAGGCTTATGTCTACTGAAGGCAATTAACGAGCAATG 1652  
 676 lIleGlnLeuIleValAlaArGArGgIleSerLySAsnGlyuLeuLySse 692  
 |||  
 1653 ATCCAGCTTATGTGTCAAGAGAGATTAAGCAAGTGCAATAGACTGAAGTC 1702  
 692 rProGlySerProGlyProGlyuLeuProIleGlyuThrAlaLeuAspA 709  
 |||  
 1703 ACCTGGAGGCCCCCTGACCTGAGCTGCCCATTTGAAGAACGCGTTGATG 1752  
 709 sPArGgLyuArGArGgIleSerHisSerLeuTySerGlyIleGlyGlyLeu 725  
 |||  
 1753 ATAGAGAACGAAAGAAATTCCTATCCCTACAGTGGGATTTGAGGGGCTT 1802  
 726 AspGlyuSerProSerArGAsnAlaAlaLeuSerArGgIleMetGlyLys 742  
 |||  
 1803 GATGAATGCCCCAGACAGAAATGCTGCTCAGTAGATGAATAGT..... 1844  
 742 rGlyLySgLyGlnLeuSerProThrValAsnMetProGlnAspAspThrV 759  
 |||  
 1845 .GGTAATATACAGCTGCCCTACAGTGAATATGCCCAAGATGACACTG 1893  
 759 aIleIleIleGlyuAspAspArGLeuProValLeuProHisLeuSerAsp 775  
 |||  
 1894 TCATTATAGAAAGATGACAGGTTGCCAGTGTCTCTCCACATCTCTCGAC 1943  
 776 GlnSerSerSerSerHisAspAspValGlyPheValThrAlaAspAl 792  
 |||  
 1944 CAGTCTCTTCCAGCTCCCATGATGATGTGGGTTTGTGAGCGCAGATGC 1993  
 792 aGlyThrTrpAlaLySAlaAlaIleSerAspSerAlaAspCysSerLeuS 809  
 |||  
 1994 TGGTACTTGGGCCAAGGCTCAATCACTGATTCAGCCGACGTGCTTTGA 2043  
 809 eTrProAspValAspProValLeuAlaPheGlnArGgLyGlyPheGlyArg 825  
 |||  
 2044 GTCACAGATGTTGATCCAGTTCCTGCTTTTCAMACAGAGGATTTGGAGCT 2093  
 826 GlnSerMetSerGlyuLySArGgThrLyGlnPheSerAspAlaSerGlnLe 842  
 |||  
 2094 CAGACT..... 2099  
 842 uAspPheValLySThrArGlySerLySserMetAspLeuGlyIleAlaAla 859  
 |||  
 2100 .....G 2100  
 859 sPGLuThrLySLeuAsnThrValAspAspGlnLySAlaGlySerProSer 875

2101 ACAGACTAACTCAATACAGTGGATGACCAAGAACGGTCTCCAGC 2150  
 876 ArgaspaValGlyProSerLeuGlyLeuLysLysSerSerSerLeuGluSe 892  
 2151 AGAGATGTGGGCTCTCCCTGGGTCTCAAGAACTCAAGCTCATTTGAGAG 2200  
 892 rleuGlnThrAlaValAlaGluValThrLeuAsnGlyAspIleProPheH 909  
 2201 TCTGCAGACCGCAGCTTCCGAGTGCATTTGAATGGGATATCTCTTCC 2250  
 909 isArgProArgProArgTylelleArgGlyArgGlyCysAsnGluSerPhe 925  
 2251 ATCGTCCAGCGCGCGGATATCAAGAGCAGGAGTGCATGAGAGGCTTC 2300  
 926 ArgAlaAlaIleAspLysSerTyrAspLysProAlaValAspAspAsp 942  
 2301 AGAGCTGCCATGCACAAATCTATGATTAACCCGGGTAGATGATGATGA 2350  
 942 pGluGlyMetGluThrLeuGluGluAspThrGluGluSerArgSerG 959  
 2351 TGAAGGCGATGGAGACCTTGGAGAGACACACAGAAATTCACAGATCAG 2400  
 959 lyArgGluSerValSerThrAlaSerAspGlnProSerHisSerLeuGlu 975  
 2401 GGAGAGAGTCTGTATCCACAGCCAGTATCAGCTCCACACTCTGGAG 2450  
 976 ArgGlnMetAsnGlyAsnGlnGluLysGlyAspLysThrAspArgLysLy 992  
 2451 AGACAAATGAATGGAACCAAGCAAGAGGTGATAGACTGATGAAAAA 2500  
 992 sAspLysThrGlyLysGlyLysLysLysAspArgAspLysGlyLysAspL 1009  
 2501 GCATTAATCTGAAAAAGAAAGAAAGATAGAGATAGAGAGAGATA 2550  
 1009 ysmetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArg 1025  
 2551 AATATGAAGCCAAAGAAAGGAAATGCTGAGAGGCTTGGGAGACATGTCAGC 2600  
 1026 PheGlyLysHisArgLysAspAspLys 1034  
 2601 CTTCCTCAACTGACGCCGAGAGAGAGA 2627

seq\_name: gb\_pr:AK024668

seq documentation\_block: 2116 bp mRNA linear PRI 29-SEP-2000  
 LOCUS AK024668  
 DEFINITION Homo sapiens cDNA: FLJ21015 f1s, clone CA05730, highly similar to  
 AF252293 Homo sapiens PAR3 (PAR3) mRNA.  
 ACCESSION AK024668  
 VERSION AK024668.1 GI:10437004  
 KEYWORDS oligo capping; f1s (full insert sequence).  
 SOURCE Homo sapiens primary endothelial cells of human coronary artery  
 cDNA to mRNA, clone\_id:CAE clone:CAE05730.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (sites)  
 Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,  
 Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,  
 Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
 NEDO human cDNA sequencing project  
 Unpublished (2000)  
 2 (bases 1 to 2116)  
 Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
 Shibahara,T., Tanaka,T. and Nakamura,Y.  
 Direct Submission  
 Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio  
 Sugano, Institute of Medical Science, University of Tokyo,  
 Laboratory of Genome Structure Analysis, Human Genome Center;  
 Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
 E-mail:cdna@aims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
 Fax:81-3-5449-5416)  
 NEDO human cDNA sequencing project supported by Ministry of

FEATURES  
 source International Trade and Industry of Japan; cDNA full insert  
 sequencing; Research Association for Biotechnology; cDNA library  
 construction, 5'- & 3'-end one pass sequencing; Department of  
 Virology and Human Genome Center, Institute of Medical Science,  
 University of Tokyo (partly supported by Science and Technology  
 Agency).  
 Location/Qualifiers  
 1..2116  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="Primary endothelial cells of human coronary  
 artery"  
 /clone="CAE05730"  
 /clone\_id="CAE"  
 /note="cloning vector pME18SFL3"  
 1..2116  
 /note="highly similar to AF252293 Homo sapiens PAR3 (PAR3)  
 mRNA"  
 BASE COUNT 651 a 476 c 571 g 418 t  
 ORIGIN  
 alignment\_scores: Length: 683  
 Quality: 3512.50 Gaps: 1  
 Ratio: 5.165  
 Percent Similarity: 99.561 Percent Identity: 99.561  
 alignment\_block:  
 US-09-757-781-2 x AK024668 ..  
 Align seg 1/1 to: AK024668 from: 1 to: 2116  
 674 GlymetIleGlnLeuIleValAlaArgArgIleSerLysCysAsnGlu 690  
 2 GGAATGATCCACGCTTATGTTGCAAGAGAGATTAAGCAAGTCATGAGCT 51  
 690 uLysSerProGlySerProProGlyProGluLeuProIleGluThrAla 707  
 52 GAAGTCACCTGGAGCCGCCCTGGAGCTGAGCTGCCATTAAGAACGCT 101  
 707 euAspAspArgGluArgArgIleSerHisSerLeuLysSerGlyIleGlu 723  
 102 TCGATGATGAGAGMAGAAATTCCTCCCTCTACAGGGGATGAG 151  
 724 GlyLeuAspGluSerProSerArgAsnAlaIleuSerArgIleMetG 740  
 152 GGGCTTGATGATCCGCCAGCAGAAATGCTGCCCTCAGTAGAGATG. 199  
 740 yGluSerGlyLysTyrGlnLeuSerProThrValAsnMetProGlnAsp 757  
 200 .....GGTAATACACAGCTGCTCCCTACAGTGAATATGCCCAAGATG 242  
 757 spThrValIleIleGluAspAspArgLeuProValLeuProHisLeu 773  
 243 ACATGTGATTAAGAAATGACAGAGTTCACAGTCTCTCCACATCTC 292  
 774 SerAspLysSerSerSerSerHisAspAspValGlyPheValThrAl 790  
 293 TCTGACAGTCTCTCTCCACCTCCATGATGATGGCTTTGTACGCG 342  
 790 aAspAlaGlyThrTyrAlaLysAlaIleSerAspSerAlaAspCys 807  
 343 AGATGCTGACTTGGGCGCAAGCTCAATCACTGATTCACCGACTGCT 392  
 807 erLeuSerProAspValAspProValLeuAlaPheGlnArgGluGlyPhe 823  
 393 CTTCGAGTCCAGATGTTGATCCAGTCTTCTTTCAACAGAGAGATT 442  
 824 GlyArgGlnSerMetSerGluLysArgThrLysGlnPheSerAspAla 840  
 443 GGACCTCAGAGTATCTCGAAAAACGCAAGCAATTTTCAGATGCGAG 492  
 840 rGlnLeuAspPheValLysThrArgLysSerLysSerMetAspLeuGly 857

```
|||||
493 TCATTTGGATTTCGTTAAACACGAAATCAAAAAGCATGGATTAGGTA 542
857 lealaaspgluThrLysLeuasnThrValAspAspGluLysAlaGlySer 873
543 TAGGTGACGAGACTAACTCAATACAGTGCATGACGAGAAAGCGATTCT 592
874 ProSerArgAspValGlyProSerLeuGlyLeuLysLysSerSerSer 890
593 CCCAGCAGAAATGTGGCTCTCCCTGGGTCTGAAGAAAGTCAACCTGCTT 642
890 uGluSerLeuGlnThrAlaValAlaGluValThrLeuasnGlyAspIleP 907
643 GGAAGCTCTGCAGACCCGACAGTGCAGGTCACTTGAATGGGATATTC 692
907 roPheHisArgProArgProArgIleIleArgGlyArgGlyCysAsnGlu 923
693 CTTTCCATCGTCCGCGCCGCGGATATCAGAGCAGGAGGATGCAAATGAG 742
924 SerPheArgAlaAlaIleAspLysSerTyrAspLysProAlaValAspAs 940
743 AGCTTCAGACTCGCATCGACAAATCTTATGATTAACCCGCGGTAGATGA 792
940 pAspAspGluGlyMetGluThrLeuGluGluAspThrGluGluSerSerA 957
793 TGAATGATGAAGGCAATGAGACCTTGAAAGAGACAGAAAGAAAGTTCAA 842
957 rGSerGlyArgGluSerValSerThrAlaSerAspGlnProSerHisSer 973
843 GATCAGGGAGAGAGTCTGTATCCACAGCCACTGATCAGCCTTCCACTCT 892
974 LeuGluArgGlnMetLsnGlyAsnGlnGluLysGlyAspLysThrAspAr 990
893 CTGAGAGACAAATGAATGCAACCAAGACAAAGAGTATAGACTGATAG 942
990 GlyLysAspLysThrGlyLysGlyLysLysLysAspArgAspLysGlu 1007
943 AAAAAGAGATAAACTGGAAGAAAGAAAGAAAGATAGATAGAGTAAGGA 992
1007 ysaAspLysMetLysAlaLysLysLysMetLeuLysGlyLeuGlyAspMet 1023
993 AGGATAAATATCAAGCCAGAGAGGAAATGCTGAAGGCTTGGAGACTG 1042
1024 PheArgPheGlyLysHisArgLysAspAspLysIleGluLysThrGly 1040
1043 TTCAGGTTTGGCAAAACATCGAAAGATGACAAAGATTGAGAAACGGGTAA 1092
1040 sIleLysIleGlnGluSerPheThrSerGluGluGluArgIleArgMetL 1057
1093 AATATAAATACAGGAATCTTTACATCAGAGAGAGAGATACGAATGA 1142
1057 ysgGlnGlnGlnLuarGileGlnAlaLysThrArgGluPheArgLuarG 1073
1143 AGCAGAGACAGAGAGATTCAGACCAAACTCGAAGATTTAGGAAACGA 1192
1074 GlnAlaArgGluArgAspTyrAlaGluIleGlnAspPheHisArgThrPh 1090
1193 CAACCTCGAGAGCCTGACTATGCTGAATTCAGAAATTTTCATCGACAT 1242
1090 eGlyCysAspAspGluLeuMetTyrGlyValSerSerTyrGluGly 1107
1243 TGGCTGTGATGATGATTAATGTAATGGGGAGTTCTTCTATGAAAGGTT 1292
1107 erMetAlaLeuAsnAlaArgProGlnSerProArgGluGlyHisMetMet 1123
1293 CCATGGCTCTCAACGCTAGACCTCAGAGCCACAGAGAGGCAATATGATG 1342
1124 AspAlaLeuTyrAlaGlnValLysLysProArgAsnSerLysProSerPr 1140
1343 GATGCTTTGTATGCCCAAGTCAAGAGCCCGGAATTCCAAACCCCTCAC 1392
1140 oValAspSerAsnArgSerThrProSerAsnHisAspArgIleGlnArgL 1157
|||||
1393 TGTAGACAGTAACAGATCACTCTAGCAATCATGATCGGATACAGCGTC 1442
1157 euArgGlnGluPheGlnGlnAlaLysGlnAspGluAspValGluAspArg 1173
1443 TGAGGCAAGAAATTTACAGCAACCAAGCAAGATGAAGATGTGAAGATCGT 1492
1174 ArgArgThrTyrSerPheGluGlnProThrProAsnAlaArgProAlaTh 1190
1493 CGCGGAGACTATATGTTTTCAGCAACCCGCGCGCAACGACGCGCGGAC 1542
1190 rGlnSerGlyArgHisSerValSerValGluValGlnMetGlnArgGlnA 1207
1543 GCAGAGCGGGCGACACTGTGTCTCGTGAGGTGCAGATGCAGCGGCGAC 1592
1207 rGlnGlnGluLuarGluSerSerGlnGlnAlaGlnArgGlnTyrSerSer 1223
1593 GCGAGAGAGAGACCGGAGAGCTCCAGCAGGGCCAGCGCATACAGCTCT 1642
1224 LeuProArgGlnSerArgLysAsnAlaSerSerValSerGlnAspSerTyr 1240
1643 CTGCTCGGCAAGCAGGAAAAATGCCAGTCGGTCTCCAGGACTCTTG 1692
1240 pGluGlnAspTyrSerProGlyGluGlyPheGlnSerAlaLysGluAsn 1257
1693 GGAGCAGAACTACTCCCTGGGGAAAGCTTCCAGAGTCCCAAAGAAACC 1742
1257 roArgTyrSerSerTyrGlnGlySerArgAsnGlyTyrLeuGlyGlyHis 1273
1743 CCAGGTACTCCAGCTACCAAGGCTTCAGGAACGCTACTGGGAGAGACT 1792
1274 GlyPheAsnAlaArgValMetLeuGluThrGlnGluLeuAspGlnGln 1290
1793 GCGTTCAACGCCAGGCTCATGTGGAACCTCAGAGCTCTTCCGCGCAGGA 1842
1290 uGlnArgArgLysGlnGlnMetLysLysGlnProProSerGluGlyP 1307
1843 ACAGAGCGGGAAGGAGCAGAGATGAAGAGAGAGCTCTTCCGAGGGGC 1892
1307 roSerAsnTyrAspSerTyrLysLysValGlnAspProSerTyrAlaPro 1323
1893 CCAGCACTATACACTCGTATAGAAGATCCAGAGACCCAGTACGCCCT 1942
1324 ProLysGlyProPheArgGlnAspValProProSerProSerGlnValAl 1340
1943 CCCAAGGGGCCCTTCCGCCAAGATGTGCCCCCTCCCTTCACAGGTTGC 1992
1340 aArgLeuAsnArgLeuGlnThrProGluLysGlyArgProPheTyrSer 1356
1993 GAGCTGAACAGACTTCAGACTCTGAGAAAGGAGGCCCTTCTATTTCC 2041
```

**THIS PAGE BLANK (USPTO)**





```

QY 121 ATSEIEVTPSVLRANMPLHVRSSDPALIGLSTSVSDSNFSESEPSRKNTWSTTAGFL 180
Db 121 ATSEIEVTPSVLRANMPLHVRSSDPALIGLSTSVSDSNFSESEPSRKNTWSTTAGFL 180
QY 181 KONTGSPPTCDCKRKDEYRSLPRTSNMWSNOFORDNARSSLASHPVWGKLEKOEODE 240
Db 181 KONTGSPPTCDCKRKDEYRSLPRTSNMWSNOFORDNARSSLASHPVWGKLEKOEODE 240
QY 241 DGTEDNSRVEPVGHADTGLEHIFPNFSLDDVMVKLEYVNDGSGPLGIHVPPSAGGRTLG 300
Db 241 DGTEDNSRVEPVGHADTGLEHIFPNFSLDDVMVKLEYVNDGSGPLGIHVPPSAGGRTLG 300
QY 301 LLVRLKEGKGAHEHNLFRENDCIYRINDGDLRNRREFQAOHMFROAMRPIIWFHVPA 360
Db 301 LLVRLKEGKGAHEHNLFRENDCIYRINDGDLRNRREFQAOHMFROAMRPIIWFHVPA 360
QY 361 ANKEDYEOSSEKKNYSSRFPSPSOYIDNRSVNSAGLITVQAPAPLNPPEQIDSHSR 420
Db 361 ANKEDYEOSSEKKNYSSRFPSPSOYIDNRSVNSAGLITVQAPAPLNPPEQIDSHSR 420
QY 421 LPHSAHSGKPPSPAPASAPONVSTTVSSGYNTKKIGKRLNIOLKKTGEGIGFITSRDY 480
Db 421 LPHSAHSGKPPSPAPASAPONVSTTVSSGYNTKKIGKRLNIOLKKTGEGIGFITSRDY 480
QY 481 TIGGSAPITYKNILPRGAALODGRLKAGDRLIEVNGVDLYGKSOEYVSLNSTKMEGTV 540
Db 481 TIGGSAPITYKNILPRGAALODGRLKAGDRLIEVNGVDLYGKSOEYVSLNSTKMEGTV 540
QY 541 SLVYFROEDAFHRELNAEPSONQIPRETKAEDIEDVILTPDGRREFLTFEVPINDSGSAG 600
Db 541 SLVYFROEDAFHRELNAEPSONQIPRETKAEDIEDVILTPDGRREFLTFEVPINDSGSAG 600
QY 601 LGVSVAGNRSKEHNAHDLGIFVKSIIINGGAASDGRRLVNDOLIAVNGESLIGKTODAME 660
Db 601 LGVSVAGNRSKEHNAHDLGIFVKSIIINGGAASDGRRLVNDOLIAVNGESLIGKTODAME 660
QY 661 TLRSNSTGCKRMQIOLIVARRISKCNELKSPGSPPELLEIALDREKRISHSLYS 720
Db 661 TLRSNSTGCKRMQIOLIVARRISKCNELKSPGSPPELLEIALDREKRISHSLYS 720
QY 721 GLEGLEDESRRNALSRINGESGKYOIPLSTMPODDVYIIEDDLPYLPHLSDQSSSS 780
Db 721 GLEGLEDESRRNALSRINGESGKYOIPLSTMPODDVYIIEDDLPYLPHLSDQSSSS 780
QY 781 SHDDVGFYADAGTWAKAAISDSADCSLSPDYDPLAFQREGFGROSKSEKRTQFSDAS 840
Db 781 SHDDVGFYADAGTWAKAAISDSADCSLSPDYDPLAFQREGFGROSKSEKRTQFSDAS 840
QY 841 QUDFVKTRKSKSMOIGIADEKTLNVDQKAGSPSRDVGPSLGLKSSLSLQTAVAEV 900
Db 841 QUDFVKTRKSKSMOIGIADEKTLNVDQKAGSPSRDVGPSLGLKSSLSLQTAVAEV 900
QY 901 TLNGDIPFHRPRPIIRGCGNESFRALIDKSYDKPAVDDDEGEMTELEETESSSSGR 960
Db 901 TLNGDIPFHRPRPIIRGCGNESFRALIDKSYDKPAVDDDEGEMTELEETESSSSGR 960
QY 961 ESVASTASPOPSHSLEROMNGNOEGKDTDKKDKTGKSKKDRKEDKMAKAKGMKGL 1020
Db 961 ESVASTASPOPSHSLEROMNGNOEGKDTDKKDKTGKSKKDRKEDKMAKAKGMKGL 1020
QY 1021 GDMERFGKHRRDKIETKTGKIKIOESFTSEEBRIKMOBERIOAKTREFREKQAREROY 1080
Db 1021 GDMERFGKHRRDKIETKTGKIKIOESFTSEEBRIKMOBERIOAKTREFREKQAREROY 1080
QY 1081 AEIODEFRTGCDDELMTGYGVSSYEGSMALNARPOSPRECHMDALYAQVKKRNKPPSP 1140
Db 1081 AEIODEFRTGCDDELMTGYGVSSYEGSMALNARPOSPRECHMDALYAQVKKRNKPPSP 1140
QY 1141 VDSNRSTPSNHDRIQRLROFOAKODEVEDERBRRTYSEFQPPNAPRPAITQSGRHSVAYE 1200
Db 1141 VDSNRSTPSNHDRIQRLROFOAKODEVEDERBRRTYSEFQPPNAPRPAITQSGRHSVAYE 1200
QY 1201 VQORORQBERESSQAOQROYSSLPROSKRNASVSQDSMEQWYSPGEGFQSAKENPRYS 1260

```

```

Db 1198 VQORORQBERESSQAOQROYSSLPROSKRNASVSQDSMEQWYSPGEGFQSAKENPRYS 1257
QY 1261 SYGSRNGYIGGIFENARVLETQELLRODRKDEQOMKOPPSGSPSYSDYKRVOPPS 1320
Db 1258 SYGSRNGYIGGIFENARVLETQELLRODRKDEQOMKOPPSGSPSYSDYKRVOPPS 1317
QY 1321 YAPPKGPFRODVPSPSQVARNRLQTPPEKGRPEFS 1356
Db 1318 YAPPKGPFRODVPSPSQVARNRLQTPPEKGRPEFS 1353

RESULT 2
ID Q96RM6 PRELIMINARY; PRT: 1273 AA.
AC Q96RM6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DI 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN LONG VARIANT
DE B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fang C., Xu Y.;
RT "Exon/Intron Structure And Splicing Variants of A Novel Human Polarity
RT Gene, hASIP.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF332593; AAK69193.1; -
SQ SEQUENCE 1273 AA; 141730 MW; 53C1A94DBC87341E CRC64;

Query Match 92.9%; Score 6537.5; DB 4; Length 1273;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 1273; Conservative 0; Mismatches 0; Indels 83; Gaps 4;

QY 1 MKTVYCGRTRRVYVPGDGMKVFSLIOQAVTRYRRAIAKDPYWTIOVHRLHGDGILD 60
Db 1 MKTVYCGRTRRVYVPGDGMKVFSLIOQAVTRYRRAIAKDPYWTIOVHRLHGDGILD 60
QY 61 LDDILDDVADKRLVAVEDEDPHHGCGDTSSGTSQSPETFGSELCTNNVSAFOPIQ 120
Db 61 LDDILDDVADKRLVAVEDEDPHHGCGDTSSGTSQSPETFGSELCTNNVSAFOPIQ 120
QY 121 ATSEIEVTPSVLRANMPLHVRSSDPALIGLSTSVSDSNFSESEPSRKNTWSTTAGFL 180
Db 121 ATSEIEVTPSVLRANMPLHVRSSDPALIGLSTSVSDSNFSESEPSRKNTWSTTAGFL 180
QY 181 KONTAGSPPTCDCKRKDEYRSLPRTSNMWSNOFORDNARSSLASHPVWGKLEKOEODE 240
Db 181 KONTAGSPPTCDCKRKDEYRSLPRTSNMWSNOFORDNARSSLASHPVWGKLEKOEODE 240
QY 241 DGTEDNSRVEPVGHADTGLEHIFPNFSLDDVMVKLEYVNDGSGPLGIHVPPSAGGRTLG 300
Db 241 DGTEDNSRVEPVGHADTGLEHIFPNFSLDDVMVKLEYVNDGSGPLGIHVPPSAGGRTLG 300
QY 301 LLVRLKEGKGAHEHNLFRENDCIYRINDGDLRNRREFQAOHMFROAMRPIIWFHVPA 360
Db 301 LLVRLKEGKGAHEHNLFRENDCIYRINDGDLRNRREFQAOHMFROAMRPIIWFHVPA 360
QY 361 ANKEDYEOSSEKKNYSSRFPSPSOYIDNRSVNSAGLITVQAPAPLNPPEQIDSHSR 420
Db 361 ANKEDYEOSSEKKNYSSRFPSPSOYIDNRSVNSAGLITVQAPAPLNPPEQIDSHSR 420
QY 421 LPHSAHSGKPPSPAPASAPONVSTTVSSGYNTKKIGKRLNIOLKKTGEGIGFITSRDY 480
Db 421 LPHSAHSGKPPSPAPASAPONVSTTVSSGYNTKKIGKRLNIOLKKTGEGIGFITSRDY 480
QY 481 TIGGSAPITYKNILPRGAALODGRLKAGDRLIEVNGVDLYGKSOEYVSLNSTKMEGTV 540

```

Db 481 TTGSAPIYVKNILPRGAIODGRLKADRLLEVNGVLVCKSOEYVSLRSTMECTV 540  
QY 541 SLVFRQEDAEHPRELNAPEQOMQIPKETKAEDIEDIVLPDGTREPLTFEYPLNDGSAG 600  
Db 541 SLVFRQEDAEHPREL-----KAEDIEDIVLPDGTREPLTFEYPLNDGSAG 587  
QY 601 LGVSVKGRSKENHADLGIFFKSIINGGAASKDGLRVNDOLIANVNGSLGKTQDAME 660  
Db 588 LGVSVKGRSKENHADLGIFFKSIINGGAASKDGLRVNDOLIANVNGSLGKTQDAME 647  
QY 661 TLRRSMSTEGNKGMIQILIVARRISKCNELKSPSGPELPIETALDRERRISHLTS 720  
Db 648 TLRRSMSTEGNKGMIQILIVARRISKCNELKSPSGPELPIETALDRERRISHLTS 707  
QY 721 GIEGLDESPSRNAALSRIINGESGKYLSPTVNMPODDTVIIEEDRLPYLPPLHSDQSSSS 780  
Db 708 GIEGLDESPSRNAALSRIIM---GKYQLSPTVNMPODDTVIIEEDRLPYLPPLHSDQSSSS 764  
QY 781 SHDVGFTYADAGTYAKAAISDSADCSLSPVDVPLAFQREGFGRO----- 840  
Db 765 SHDVGFTYADAGTYAKAAISDSADCSLSPVDVPLAFQREGFGRO----- 810  
QY 841 QLDFVTKRKSMDLGIIDETKRLNTVDDQKAGSPSRDYGPSLGLKSSSLESLOTAVAEV 900  
Db 811 -----IADETKRLNTVDDQKAGSPSRDYGPSLGLKSSSLESLOTAVAEV 854  
QY 901 TLNGDIPHRPRRIIRGRCNESFRRAIDRSYDKPAVDDEGMEETLEEDTESSRSGR 960  
Db 855 TLNGDIPHRPRRIIRGRCNESFRRAIDRSYDKPAVDDEGMEETLEEDTESSRSGR 914  
QY 961 ESVSASQOPSHSLERONGNOEKDRTDRKDKTGKCKKADROKEMKAKKMKLGL 1020  
Db 915 ESVSASQOPSHSLERONGNOEKDRTDRKDKTGKCKKADROKEMKAKKMKLGL 974  
QY 1021 GDMFRGKHRRDKLETKGIKIQESFTSEERIIMKQEOERIQAKTRFRRQARNDY 1080  
Db 975 GDMF-----RIQAKTRFRRQARNDY 997  
QY 1081 AEIODFHRTFGCDDLEMTAGVSSYEGSMALNARPOSPRGHMDALYAQVKPRNSKPS 1140  
Db 998 AEIODFHRTFGCDDLEMTAGVSSYEGSMALNARPOSPRGHMDALYAQVKPRNSKPS 1057  
QY 1141 VDSNSTPSNDRIQRLKQEQAKODDEDVEDRRRTYSFEQWPNARPATOGGRHSYVE 1200  
Db 1058 VDSNSTPSNDRIQRLKQEQAKODDEDVEDRRRTYSFEQWPNARPATOGGRHSYVE 1117  
QY 1201 VQMQROEROERESSQOARQYSSILPROGRKNASYSODSWEQNTSPGEGFQSAKENPRYS 1260  
Db 1118 VQMQROEROERESSQOARQYSSILPROGRKNASYSODSWEQNTSPGEGFQSAKENPRYS 1177  
QY 1261 SYQGRNGYLGGHGFANAVMETQELRLQEOQRKEQOMKQPPSEGPSNYDSYKRVQDS 1320  
Db 1178 SYQGRNGYLGGHGFANAVMETQELRLQEOQRKEQOMKQPPSEGPSNYDSYKRVQDS 1237  
QY 1321 YAPPKGPRQDVPSPSOVARLNRLQTEPKGRPYS 1356  
Db 1238 YAPPKGPRQDVPSPSOVARLNRLQTEPKGRPYS 1273

## RESULT 3

ID 09NYE6 PRELIMINARY: PRT: 1266 AA.  
AC 09NYE6: 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
GN PAR3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eultheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RX MEDLINE=20394296; PubMed=10934474;  
RA Jobery G., Petersen C., Gao L., Macara I.G.;  
RT "The cell-polarity protein Par6 links Par3 and atypical protein kinase  
C to Cdc42.";  
RL Nat. Cell Biol. 2:531-539(2000).  
DR EMBL: AF252293; AAF71530.1; -.  
DR HSSP: Q12923; 3PD2.  
DR InterPro: IPR001478; PDZ.  
DR Pfam: PF00595; PDZ; 3.  
DR SMART: SM00228; PDZ; 3.  
DR PROSITE: PS0106; PDZ; 3.  
SQ SEQUENCE 1266 AA; 141071 MW; 92DE51B68081A4A2 CRC64;

Query Match 92.2%; Score 6485; DB 4; Length 1266;  
Best Local Similarity 93.4%; Pred. No. 0;  
Matches 1266; Conservative 0; Mismatches 0; Indels 90; Gaps 4;

QY 1 MKVTVCFGRTRVVPCGDHMKVFSLIQAVTRYKRAIAKDPNWIQVHRLHGDGILD 60  
Db 1 MKVTVCFGRTRVVPCGDHMKVFSLIQAVTRYKRAIAKDPNWIQVHRLHGDGILD 60  
QY 61 LDDILCDVADKDRLYAVFDEQDPHHGDPGTSASTGTOSPEIFGSELCTNNVSAFQPYQ 120  
Db 61 LDDILCDVADKDRLYAVFDEQDPHHGDPGTSASTGTOSPEIFGSELCTNNVSAFQPYQ 120  
QY 121 ATSEIEVTPSVLRANPPLVHRSSDPALIGLSTSVSDSNFSSEPRKPTMTSTAGFL 180  
Db 121 ATSEIEVTPSVLRANPPLVHRSSDPALIGLSTSVSDSNFSSEPRKPTMTSTAGFL 180  
QY 181 KONTAGSPKTCDDKDKDENYRSLPRDTSNWSNOFORNARSSLSASHPMVKLEKQED 240  
Db 181 KONTAGSPKTCDDKDKDENYRSLPRDTSNWSNOFORNARSSLSASHPMVKLEKQED 240  
QY 241 DGTEDNSRVEYVGAADTGLEHINPFSLDDMYKLYEVPNDGPGFLGHVVPFSAARGRTLG 300  
Db 197 DGTEDNSRVEYVGAADTGLEHINPFSLDDMYKLYEVPNDGPGFLGHVVPFSAARGRTLG 256  
QY 301 LLYKRLKGGKAEHNLRFENDCIYRINDGDLRNRFEQAOHFRAMRPTIIMFHVPA 360  
Db 257 LLYKRLKGGKAEHNLRFENDCIYRINDGDLRNRFEQAOHFRAMRPTIIMFHVPA 316  
QY 361 ANKEOYEQLQSEKKNYSSRSPSQYIDNRSVNSAGLHTVQARLNLHPPEQIDSHSR 420  
Db 317 ANKEOYEQLQSEKKNYSSRSPSQYIDNRSVNSAGLHTVQARLNLHPPEQIDSHSR 376  
QY 421 LPHSAHPGSKPPSAPASAFQNVFTTVSSGYNTKTIQKRLNIOCLKGTGLGFSITSRDY 480  
Db 377 LPHSAHPGSKPPSAPASAFQNVFTTVSSGYNTKTIQKRLNIOCLKGTGLGFSITSRDY 436  
QY 481 TTGSAPIYVKNILPRGAIODGRLKADRLLEVNGVDLVGKSOEYVSLRSTMECTV 540  
Db 437 TTGSAPIYVKNILPRGAIODGRLKADRLLEVNGVDLVGKSOEYVSLRSTMECTV 496  
QY 541 SLVFRQEDAEHPRELNAPEQOMQIPKETKAEDIEDIVLPDGTREPLTFEYPLNDGSAG 600  
Db 497 SLVFRQEDAEHPREL-----KAEDIEDIVLPDGTREPLTFEYPLNDGSAG 543  
QY 601 LGVSVKGRSKENHADLGIFFKSIINGGAASKDGLRVNDOLIANVNGSLGKTQDAME 660  
Db 544 LGVSVKGRSKENHADLGIFFKSIINGGAASKDGLRVNDOLIANVNGSLGKTQDAME 603  
QY 661 TLRRSMSTEGNKGMIQILIVARRISKCNELKSPSGPELPIETALDRERRISHLTS 720  
Db 604 TLRRSMSTEGNKGMIQILIVARRISKCNELKSPSGPELPIETALDRERRISHLTS 663  
QY 721 GIEGLDESPSRNAALSRIINGESGKYLSPTVNMPODDTVIIEEDRLPYLPPLHSDQSSSS 780  
Db 664 GIEGLDESPSRNAALSRIIM---GKYQLSPTVNMPODDTVIIEEDRLPYLPPLHSDQSSSS 720  
QY 781 SHDVGFTYADAGTYAKAAISDSADCSLSPVDVPLAFQREGFGROSMSEKRTKQPSDAS 840

	Mismatches	63;	Indels	19; Gaps	2
	Conservative	63;			
Matches	1216;				
	TCGAAATRYRRRAIAKDPNWTQVHREHGSGILD	60			

[illegible]

QY 1081 AEIODEFHRTFGCDDDELMYGVSSYEGSMALNARPOS PREGHMMDALYAQVKKRNSKPS 1140  
Db 1081 AEIODEFHRTFGCDDDELMYGVSSYEGSMALNARPOS PREGHMMDALYAQVKKRNSKPS - 1139  
QY 1141 VDSNSTSTSNHRIQRLRQEFQAKODEVEYRRRTTYSFEQWPNARPAQSGRHSVSE 1200  
Db 1140 -DSNSTSTSNHRIQRLRQEFQAKODEVEYRRRTTYSFEQWPNARPAQSGRHSVSE 1198  
QY 1201 VOMOROROEERSSQOARQOYSSLPROSKRNKNSVSDSDWEQNTSPGEGFQSAKENPYS 1260  
Db 1199 VQVQROQOEERSSFOAQROYSLSLPROSKRNKNSVSDSDWEQNTSPGEGFQSAKENPYS 1258  
QY 1261 SYQSRNGYLGSHGFNARVMELOELRQOEORRQEQOMKQPPSEGPSNYDSYKKVODPS 1320  
Db 1259 SYQSRNGYLGSHGFNARVMELOELRQOEORRQEQOLKQPPADG----- 1304  
QY 1321 YAPKGPQODVPSPSOVARLNRLQTEPEKGRFYS 1356  
Db 1305 --YKGPQODVPSPSOVARLNRLQTEPEKGRFYS 1337  
RESULT 5  
Q99NH2 PRELIMINARY: PRT: 1333 AA.  
AC Q99NH2: 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PAR-3 180 KDA ISOFORM.  
GN PARD3 OR PAR3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NIH SWISS;  
RX MEDLINE=9912117; PubMed=9920925;  
RA Lin D., Gish G.D., Songyang Z., Pawson T.;  
RT "The carboxyl terminus of B class ephrins constitutes a PDZ domain  
blinding motif.";  
RL J. Biol. Chem. 274:3726-3733(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NIH SWISS;  
RX MEDLINE=20394297; PubMed=10934475;  
RA Lin D., Edwards A.S., Fawcett J.P., Mhamali G., Scott J.D., Pawson T.;  
RT "A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and aPKC  
signaling and cell polarity.";  
RL Nat. Cell Biol. 2:540-547(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NIH SWISS;  
RA Lin D.C. Jr., Pawson T.J.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY026057; AAK07669.1; -  
DR HSSP; Q12923; 3PDZ.  
DR MGD; MGI:2135608; Pard3.  
DR InterPro; IPR001478; PDZ.  
DR Pfam; PF00595; PDZ; 3.  
DR SMART; SM00228; PDZ; 3.  
DR PROSITE; PS0106; PDZ; 3.  
SQ SEQUENCE 1333 AA; 149060 MW; AF67825C66DCE86 CRC64;

Query Match 89.7%; Score 6310.5; DB 11; Length 1333;  
Best Local Similarity 89.7%; Pred. No. 0;  
Matches 1216; Conservative 60; Mismatches 57; Indels 23; Gaps 4;

QY 1 MKVTVCFGFTRVYVPCGCGHMKVFSLIQAAVTRRYRAIKAKDPNMYIQVRLHEHGDGILD 60  
Db 1 MKVTVCFGFTRVYVPCGCGHMKVFSLIQAAVTRRYRAIKAKDPNMYIQVRLHEHGDGILD 60

QY 61 LDDILCDVADDDKRLVAVEDEQDPHHGCDGTSASTGTQSPETFGSELGNTNNVSAFQPYQ 120  
Db 61 LDDILCDVADDDKRLVAVEDEQDPHHGCDGTSASTGTQSPETFGSELGNTNNVSAFQPYQ 120  
QY 121 ATSEIEVTPSVLARNPPLHVRRSSDPALIGLSTVSDSNFSSEEPKRNPTRWSTTAGEL 180  
Db 121 ATSEIEVTPSVLARNPPLHVRRSSDPALIGLSTVSDSNFSSEEPKRNPTRWSTTAGEL 180  
QY 181 KONFAGSPPTCDKRNKENTRSLPRDTSNMSNOPORANASSLSASPMYKXMLEKEQDE 240  
Db 181 KONFAGSPPTCDKRNKENTRSLPRDTSNMSNOPORANASSLSASPMYKXMLEKEQDE 240  
QY 241 DGEEDNSRVEEYVGHADTLEHINPFSLDDMYVLVEVPNDGFGIGHVVPFSAFGRTTG 300  
Db 241 EGEEDNSRVEEYVGHADTLEHINPFSLDDMYVLVEVPNDGFGIGHVVPFSAFGRTTG 300  
QY 301 LLYKLEKGGKAEHNLFEHNDCTVRIINDGDLNRRFEQAQHFROAMKRTPIYFHVPA 360  
Db 301 LLYKLEKGGKAEHNLFEHNDCTVRIINDGDLNRRFEQAQHFROAMKRTPIYFHVPA 360  
QY 361 ANKEOYEOLSOSEKKNYYSRFPDSOYIDNRSVNSAGLHTVQRAVRLNHPQIDSHSR 420  
Db 361 ANKEOYEOLSOSEKKNYYSRFPDSOYIDNRSVNSAGLHTVQRAVRLNHPQIDSHSR 420  
QY 421 LPHSAPSGKPPSAPASAPQWNESTVSSQYNTKKIGKRLNLOLKKTEGLGFSITSRDY 480  
Db 421 LPHSAPSGKPPSAPASAPQWNESTVSSQYNTKKIGKRLNLOLKKTEGLGFSITSRDY 480  
QY 481 TIGGSAPITYKNTLIPGAAIIOGRLKAGDRLIEVNCVDLVGKSQEEVYSILSTKKEGY 540  
Db 481 TIGGSAPITYKNTLIPGAAIIOGRLKAGDRLIEVNCVDLVGKSQEEVYSILSTKKEGY 540  
QY 541 SLVVFQEDAFPHRELNAEBSQMOIPKETAEDEDVILVPDGRREFLTFEVPILNDGSGAG 600  
Db 541 SLVVFQEDAFPHRELNAEBSQMOIPKETAEDEDVILVPDGRREFLTFEVPILNDGSGAG 600  
QY 601 LGVSVKGNRSKENHADLIGIFYKSIINGAASKDRRLRVNDQILAVNGESILGKTNDAME 660  
Db 601 LGVSVKGNRSKENHADLIGIFYKSIINGAASKDRRLRVNDQILAVNGESILGKTNDAME 660  
QY 661 TLRRSSTEGCKRMGIQIYARRISKCNELKSGSPGPELPLETALDDRRRISHSLS 720  
Db 661 TLRRSSTEGCKRMGIQIYARRISKCNELKSGSPGPELPLETALDDRRRISHSLS 720  
QY 721 GIEGLDESPRNAAALSRIINGESGYOLSPVNNPQDVTYIETEDRLPYLPHILSDSSSS 780  
Db 721 GIEGLDESPRNAAALSRIINGESGYOLSPVNNPQDVTYIETEDRLPYLPHILSDSSSS 780  
QY 781 SHDVGFTVADAGTAAKAAISDSADCSLSPDVYPVLAFOREGFGROSMSSEKRTKQFSDAS 840  
Db 781 SHDVGFTVADAGTAAKAAISDSADCSLSPDVYPVLAFOREGFGROSMSSEKRTKQFSDAS 840  
QY 841 QLDFVYTRKSKSMDLGIADETKLTNYDDQAGSPSDVPSGLCKKSSLSLSIQTAVAEY 900  
Db 841 QLDFVYTRKSKSMDLGIADETKLTNYDDQAGSPSDVPSGLCKKSSLSLSIQTAVAEY 900  
QY 901 TLNGDIPFHRPRRIIRGRCNESPRAAIDKSVKDAVDDDDGMELEEDTESSSGSR 960  
Db 901 TLNGDIPFHRPRRIIRGRCNESPRAAIDKSVKDAVDDDDGMELEEDTESSSGSR 960  
QY 961 ESYVSTASDQPSHSLEROMNGNOEKDTRKDKDTEKKEKKAKDKKAKKAKGMLKGL 1020  
Db 961 ESYVSTASDQPSHSLEROMNGNOEKDTRKDKDTEKKEKKAKDKKAKKAKGMLKGL 1020  
QY 1017 GMEFRFGKHRRKDDKMEKMRKIQDSFTSEEDYVRMKEQERIOAKTRREFRERQARERDY 1076  
Db 1017 GMEFRFGKHRRKDDKMEKMRKIQDSFTSEEDYVRMKEQERIOAKTRREFRERQARERDY 1076  
QY 1081 AEIODEFHRTFGCDDDELMYGVSSYEGSMALNARPOS PREGHMMDALYAQVKKRNSKPS 1140  
Db 1077 AEIODEFHRTFGCDDDELMYGVSSYEGSMALNARPOS PREGHMMDALYAQVKKRNSKPS - 1135

QY 1141 VDSNRSTPSNHDRIORLROEFOAKODEVEDRRRTYSFEQPPNAPRATOGSRHSVYE 1200  
 Db 1136 -DSNRSTPSNHDRIORLROEFOAKODEVEDRRRTYSFEQSSSRPASOSGRHSVYE 1194  
 QY 1201 VOMOROEERESQOARQYSSLPORSRNASSVSQDSMEQONYSFGEFOAKENPYS 1260  
 Db 1195 VOMOROEERESQOARQYSSLPORSRNASSVSQDSMEQONYSFGEFOAKENPYS 1254  
 QY 1261 SYGSRNGYLGHGFFNARVLETQELLROEORKEQOMKOPSPSGPNSYDKKVVODPS 1320  
 Db 1255 SYGSRNGYLGHGFFNARVLETQELLROEORKEQOMKOPSPSGPNSYDKKVVODPS 1300  
 QY 1321 YAPKGFRODVPSPSOVARLNRLQTPKGRPFYS 1356  
 Db 1301 ---VGRPFRODVPSPSOVARLNRLQTPKGRPFYS 1333

RESULT 6  
 ID Q9BY57 PRELIMINARY; PRT: 1031 AA.  
 AC Q9BY57;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN SHORT VARIANT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN 11;  
 RP SEQUENCE FROM N.A.  
 RA Fang C., Xu Y.;  
 RT "Exon/Intron Structure and Splicing Variants of a Novel Human Polarity  
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF196186; AAK27892.1; -  
 DR HSSP; Q12923; 3pDZ.  
 DR InterPro; IPR001478; PDZ.  
 DR Pfam; PF00595; PDZ; 3  
 DR SMART; SM00228; PDZ; 3.  
 DR PROSITE; PS0106; PDZ; 3.  
 SO SEQUENCE 1031 AA; 113418 MW; 40DD593636EAB999 CRC64;

Query Match 74.9%; Score 5268.5; DB 4; Length 1031;  
 Best Local Similarity 98.8%; Pred. No. 2e-290;  
 Matches 1022; Conservative 3; Mismatches 6; Indels 3; Gaps 1;

QY 1 MKVTYCFGRTYVPCGCGHKVFSLLQOAVTRYKAIADPNWIOVHLEHGDSILD 60  
 Db 1 MKVTYCFGRTYVPCGCGHKVFSLLQOAVTRYKAIADPNWIOVHLEHGDSILD 60  
 QY 61 LDDILCDVADDKRLVAVFEDODPHGGDGTSSSTQSPETFGSELGTNNYSAFQPY 120  
 Db 61 LDDILCDVADDKRLVAVFEDODPHGGDGTSSSTQSPETFGSELGTNNYSAFQPY 120  
 QY 121 ATSELETPSVTLRANMPLHVRSSDPALIGLSTVSQSNSSSESPKRPWTSTAGFL 180  
 Db 121 ATSELETPSVTLRANMPLHVRSSDPALIGLSTVSQSNSSSESPKRPWTSTAGFL 180  
 QY 121 ATSELETPSVTLRANMPLHVRSSDPALIGLSTVSQSNSSSESPKRPWTSTAGFL 180  
 Db 121 ATSELETPSVTLRANMPLHVRSSDPALIGLSTVSQSNSSSESPKRPWTSTAGFL 180  
 QY 181 KONTAGSPKTCDDKRDENYSLPRDTSNMSNOFORNANASSLSASHPMVKMLERQODE 240  
 Db 181 KONTAGSPKTCDDKRDENYSLPRDTSNMSNOFORNANASSLSASHPMVKMLERQODE 240  
 QY 181 KONTAGSPKTCDDKRDENYSLPRDTSNMSNOFORNANASSLSASHPMVKMLERQODE 240  
 Db 181 KONTAGSPKTCDDKRDENYSLPRDTSNMSNOFORNANASSLSASHPMVKMLERQODE 240  
 QY 241 DGTEDNSRVEPGHADTGLEHINPNSLSDMYKIVEPNDGGLGIHVPPSARGRLTG 300  
 Db 241 DGTEDNSRVEPGHADTGLEHINPNSLSDMYKIVEPNDGGLGIHVPPSARGRLTG 300  
 QY 301 LTVKRLKGGKAEHNTLFRENCIYRINDGDLRNRFFQAOHMFQOARTPIIMFHVPA 360  
 Db 301 LTVKRLKGGKAEHNTLFRENCIYRINDGDLRNRFFQAOHMFQOARTPIIMFHVPA 360  
 QY 361 ANKEQYEQLSQSEKNNYSSRPSDQYIDNRSVNSAGLHTVQAPRLNHPEDIDSHSR 420

Db 361 ANKEQYEQLSQSEKNNYSSRPSDQYIDNRSVNSAGLHTVQAPRLNHPEDIDSHSR 420  
 QY 421 LPHSAHPGPPAPASAPQNFSTTVSSGYTKTKIGRLNIQLKKGEGFGFTSRDY 480  
 Db 421 LPHSAHPGPPAPASAPQNFSTTVSSGYTKTKIGRLNIQLKKGEGFGFTSRDY 480  
 QY 481 TIGGSAPYIVKNIILPRGAALIDGRILKAGDRILEVNGVLYKSGSEEVYSLRSTKMGTV 540  
 Db 481 TIGGSAPYIVKNIILPRGAALIDGRILKAGDRILEVNGVLYKSGSEEVYSLRSTKMGTV 540  
 QY 541 SLVFRQEDAFPRRLNPSOMQIPKETAEDDIYLTDPGTRFLEFVPLNDGSAG 600  
 Db 541 SLVFRQEDAFPRRLNPSOMQIPKETAEDDIYLTDPGTRFLEFVPLNDGSAG 600  
 QY 601 LGSVYKGRSKENHADIGIFVKSIIINGAASKODRLRVNDQLIANGESILKTNODAME 660  
 Db 601 LGSVYKGRSKENHADIGIFVKSIIINGAASKODRLRVNDQLIANGESILKTNODAME 660  
 QY 661 TLRRSMTEGKRGMIQIIVARRISKCNELKSPGSPGPELPIETALDRERRISHLYS 720  
 Db 661 TLRRSMTEGKRGMIQIIVARRISKCNELKSPGSPGPELPIETALDRERRISHLYS 720  
 QY 721 GIEGLDPSPSRNALSRIMGESGYQLSPVYVNPPODDYIITEDDLPLVPLPHLSQSSSS 780  
 Db 721 GIEGLDPSPSRNALSRIMGESGYQLSPVYVNPPODDYIITEDDLPLVPLPHLSQSSSS 777  
 QY 781 SHDVGFEVTADGTWAKAISDSADCSLSPDYVPIAFQREGFGROSSEKRTKOFSDAS 840  
 Db 778 SHDVGFEVTADGTWAKAISDSADCSLSPDYVPIAFQREGFGROSSEKRTKOFSDAS 837  
 QY 841 QLDPEVTRKSKMDIGIADETKLTAVDDQKAGSPRDVPSLGLTKSSLSLQTAVAEV 900  
 Db 838 QLDPEVTRKSKMDIGIADETKLTAVDDQKAGSPRDVPSLGLTKSSLSLQTAVAEV 897  
 QY 901 TLNGDIPFHRPRPRIIRGCGNESFRAIDSYKPAVDDDEGEMTLEBDTESSSGR 960  
 Db 898 TLNGDIPFHRPRPRIIRGCGNESFRAIDSYKPAVDDDEGEMTLEBDTESSSGR 957  
 QY 961 ESVSTASDQPSHSLEKONGNKGDKTDKDKTKGKKDKDKKKAKKMKLKL 1020  
 Db 958 ESVSTASDQPSHSLEKONGNKGDKTDKDKTKGKKDKDKKKAKKMKLKL 1017  
 QY 1021 GDMFRFGHRRDK 1034  
 Db 1018 GDMFSLAKLKEKR 1031

RESULT 7  
 ID Q96RM7 PRELIMINARY; PRT: 988 AA.  
 AC Q96RM7;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN SHORT VARIANT  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN 11;  
 RP SEQUENCE FROM N.A.  
 RA Fang C., Xu Y.;  
 RT "Exon/Intron Structure And Splicing Variants of A Novel Human Polarity  
 RL Submitted (Dec-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF332592; AAK69192.1; -  
 SO SEQUENCE 988 AA; 108545 MW; 89F2139B096F7F7F CRC64;

Query Match 71.3%; Score 5014; DB 4; Length 988;  
 Best Local Similarity 94.7%; Pred. No. 5.2e-276;

Matches 979: Conservative 3; Mismatches 6; Indels 46; Gaps 3;				
Qy	1	MKVTVCGRTFRVVPVPCGDMHKVPSLIDQAVTRYRKALAKDPNMYIQVHLEHGDGLID	60	
Db	1	MKVTVCGRTFRVVPVPCGDMHKVPSLIDQAVTRYRKALAKDPNMYIQVHLEHGDGLID	60	
Qy	61	LDLIDCDVADKDLVAVFDEDDPHGGDGTASSTGTQSPETFGSELGTTNNVSAFQPYQ	120	
Db	61	LDLIDCDVADKDLVAVFDEDDPHGGDGTASSTGTQSPETFGSELGTTNNVSAFQPYQ	120	
Qy	121	ATSEIEVPSVLRAMPVHVRSSDPALIGSTSVSDSNFSESESRKNPRTMSTTAGFL	180	
Db	121	ATSEIEVPSVLRAMPVHVRSSDPALIGSTSVSDSNFSESESRKNPRTMSTTAGFL	180	
Qy	181	KONTAGSPKTCDDKKDEYRSLPRDTSNMWSNQFQDNARSSLASHPVGVKLEFQODE	240	
Db	181	KONTAGSPKTCDDKKDEYRSLPRDTSNMWSNQFQDNARSSLASHPVGVKLEFQODE	240	
Qy	241	DGTEEDNSRVEPVGHADTGLEHITPNFSLDDMYKLYEVPNDGGPLGIIHVVPESARGRTLG	300	
Db	241	DGTEEDNSRVEPVGHADTGLEHITPNFSLDDMYKLYEVPNDGGPLGIIHVVPESARGRTLG	300	
Qy	301	LLVRLKLGKAEHENTLRENDCTVRINDGDLRNRFEQAQHMFRQAMRTPIIMHVVPA	360	
Db	301	LLVRLKLGKAEHENTLRENDCTVRINDGDLRNRFEQAQHMFRQAMRTPIIMHVVPA	360	
Qy	361	ANKEQYEQLSQSEKNNYSRFPSPDQYIDNRVNSAGLHTVQAPRLNHPPEQIDSHSR	420	
Db	361	ANKEQYEQLSQSEKNNYSRFPSPDQYIDNRVNSAGLHTVQAPRLNHPPEQIDSHSR	420	
Qy	421	LPHSAHPGKPPSPAPASAPQVFTTSSGYNTKKIGKRLNIOLKKGTEGLFSTSRDY	480	
Db	421	LPHSAHPGKPPSPAPASAPQVFTTSSGYNTKKIGKRLNIOLKKGTEGLFSTSRDY	480	
Qy	481	TTGSGAPLYVNNILPRGAIDODGLKAGDRLIEVNGVDLVGKSOEYVSLRSTMEGV	540	
Db	481	TTGSGAPLYVNNILPRGAIDODGLKAGDRLIEVNGVDLVGKSOEYVSLRSTMEGV	540	
Qy	541	SLLVFRODAPRRLNAEPESOMQIPKETEKADEEDIVLTPDGTREFLTFEYVPLNDGSGAG	600	
Db	541	SLLVFRODAPRRLNAEPESOMQIPKETEKADEEDIVLTPDGTREFLTFEYVPLNDGSGAG	600	
Qy	601	LGVSYKGRSKENHADLGIYFKSIINGGAASFDGLRYVNDQIIAANGESLIGKTODAME	660	
Db	601	LGVSYKGRSKENHADLGIYFKSIINGGAASFDGLRYVNDQIIAANGESLIGKTODAME	660	
Qy	661	TLRSMSTEGKRGMIQIIVARISKCNELKSPGSPGPPELPIETALDDREKRISHSLYS	720	
Db	661	TLRSMSTEGKRGMIQIIVARISKCNELKSPGSPGPPELPIETALDDREKRISHSLYS	720	
Qy	721	GIEGLDESPSRNALSRIMGESGYOLSPVTVMPODDTVIIEDDRPLVLPRLSDQSSSS	780	
Db	721	GIEGLDESPSRNALSRIMGESGYOLSPVTVMPODDTVIIEDDRPLVLPRLSDQSSSS	780	
Qy	781	SHDVGFTTAAGTAAKAAISDSADCSLSPVDVPLAFQREGFGQSMSEKRTKQPSDAS	840	
Db	781	SHDVGFTTAAGTAAKAAISDSADCSLSPVDVPLAFQREGFGQSMSEKRTKQPSDAS	840	
Qy	841	QLDFKTKSKSMQDGIADETFKLNTVDDOKAGSPRDVGPISGLKSSSLJESLQTAFAVEV	900	
Db	841	QLDFKTKSKSMQDGIADETFKLNTVDDOKAGSPRDVGPISGLKSSSLJESLQTAFAVEV	900	
Qy	901	TLNGDIPHRPRPRLIRRGCSNESFRRAIDKSYDKPAVDDDEGMETLEEDTEESRSRGR	960	
Db	901	TLNGDIPHRPRPRLIRRGCSNESFRRAIDKSYDKPAVDDDEGMETLEEDTEESRSRGR	960	
Qy	961	ESVSTAQSPSHSLERQANGNOEKDTRDKDKTGKEKKKRDKEKKDKMKAKKGMKLKL	1020	
Db	961	ESVSTAQSPSHSLERQANGNOEKDTRDKDKTGKEKKKRDKEKKDKMKAKKGMKLKL	1020	
Qy	1021	GDMFFGKHKRDKD 1034		
Db	1021	GDMFFGKHKRDKD 1034		
Qy	975	GDMFSLAKLKEPKR 988		
Db	975	GDMFSLAKLKEPKR 988		

RESULT	8	
Q96K28	PRELIMINARY:	PRF: 865 AA.
AC	Q96K28:	01-DEC-2001 (TREMBLrel. 19, Created)
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	CNDA FLJ14829 FIS, CLONE OVARC1000945, MODERATELY SIMILAR TO	
DE	RATBUS NORVEGICUS ATYPICAL PKC SPECIFIC BINDING PROTEIN.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX	NCBI_Taxid=9606;	
RN	[1]	
RP	SEQUENCE FROM N. A.	
RC	TISSUE=OVARIAN CARCINOMA;	
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,	
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,	
RA	Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,	
RA	Takehashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,	
RA	Watanabe S., Kimura K., Murekami K., Ishi S., Kawai Y., Saito K.,	
RA	Yanamoto J., Makamatsu A., Nakamura Y., Nagahari K., Masuno Y.,	
RT	Niomiya K., Iwayanagi T.;	
RT	"NEO human cDNA sequencing project."	
RL	Submitted (May-2001) to the EMBL/Genbank/DBJ databases.	
DR	EMBL: AK027735; BAB5330.1; -	
SQ	SEQUENCE 865 AA: 95161 MW; 5378BBCD406D0835 CRC64;	
Query Match 62.0%; Score 4361; DB 4; Length 865;		
Best Local Similarity 94.8%; Pred. No. 4,5e-239;		
Matches 852; Conservative 5; Mismatches 8; Indels 34; Gaps 2;		
Qy	136	MLVHRRSSDPLLIGLSTVSDSNFSSEPSRKNPRTMSTTAGFLKONTAGSPKTCDDRK 195
Db	1	MLVHRRSSDPLLIGLSTVSDSNFSSEPSRKNPRTMSTTAGFLKONTAGSPKTCDDRK 60
Qy	196	DENYRSLPDDTSNMWSNQFQDNARSSLASHPVGVKLEFQODEDGEDNSRVEPVGH 255
Db	61	DENYRSLPDDTSNMWSNQFQDNARSSLASHPVGVKLEFQODEDGEDNSRVEPVGH 120
Qy	256	ADTGLEHITPNFSLDDMYKLYEVPNDGGPLGIHVVPESARGRTLGILVRLKLGKAEHE 315
Db	121	ADTGLEHITPNFSLDDMYKLYEVPNDGGPLGIHVVPESARGRTLGILVRLKLGKAEHE 180
Qy	316	NLFRENDCTVRINDGDLRNRFEQAQHMFRQAMRTPIIMHVPAANKQYEQIDLSQSEKN 375
Db	181	NLFRENDCTVRINDGDLRNRFEQAQHMFRQAMRTPIIMHVPAANKQYEQIDLSQSEKN 240
Qy	376	NYSSRFPSPDQYIDNRVNSAGLHTVQAPRLNHPPEQIDSHSRILPHSAHPGKPPSPAP 435
Db	241	NYSSRFPSPDQYIDNRVNSAGLHTVQAPRLNHPPEQIDSHSRILPHSAHPGKPPSPAP 300
Qy	436	ASAPQVFTTSSGYNTKKIGKRLNIOLKKGTEGLFSTSRDYTTGGSAPIYVNNILP 495
Db	301	ASAPQVFTTSSGYNTKKIGKRLNIOLKKGTEGLFSTSRDYTTGGSAPIYVNNILP 360
Qy	496	RGAAIODGRLKAGDRLIEVNGVDLVGKSOEYVSLRSTKMGSTVSLVFRQEDAFHPRE 555
Db	361	RGAAIODGRLKAGDRLIEVNGVDLVGKSOEYVSLRSTKMGSTVSLVFRQEDAFHPRE 420
Qy	556	LNAEPSONOIPKETADEEDIVLTPDGTREFLTFEYVPLNDGSGAGGVYKGRSKENNA 615
Db	421	LNAEPSONOIPKETADEEDIVLTPDGTREFLTFEYVPLNDGSGAGGVYKGRSKENNA 480
Qy	616	DGIFVKSIIINGGAASKDGLRLVNDQIIAANGESLIGKTODAMETLRSMSTEGKRG 675
Db	481	DGIFVKSIIINGGAASKDGLRLVNDQIIAANGESLIGKTODAMETLRSMSTEGKRG 540
Qy	676	IOLIVARRISKCNELKSPSPGPPELPIETALDDREKRISHSLYSIGLDESPSRNAL 735
Db		IOLIVARRISKCNELKSPSPGPPELPIETALDDREKRISHSLYSIGLDESPSRNAL 735



```

Db 541 IQLIVARRISKCNELKSPGSPPELPETALDRERRRISHSLYSGIBSDSPRNAL 600
QY 736 SRIMGESGKYQLSPVNMPODDVYIIEDDRPLVPLPHLSDQSSSSSHDDVGFYADAGTW 795
Db 601 SRIM---GKQYLSPTVNMPODDVYIIEDDRPLVPLPHLSDQSSSSSHDDVGFYADAGTW 657
QY 796 AKAAISDSADCSLSPVDVYLAFOREGFGROSMSEKRTQFSDASQDLFVTRKSKMDL 855
Db 658 AKAAISDSADCSLSPVDVYLAFOREGFGRO----- 689
QY 856 GIADETKLNTVDDOKAGSPSRDVGPSLGLKSSLSLESLQTAVAEYTLNGDIPFHRPRRI 915
Db 690 ---DETKLNTVDDOKAGSPSRDVGPSLGLKSSLSLESLQTAVAEYTLNGDIPFHRPRRI 746
QY 916 IGRGNCNESFRAIDKSYDKPAVDDDDGEMTLEEDTEESSRSGRESVSTASQPSHLE 975
Db 747 IGRGNCNESFRAIDKSYDKPAVDDDDGEMTLEEDTEESSRSGRESVSTASQPSHLE 806
QY 976 ROMNGNOERKGDTRKRRKKTGKKKKDRKKKKKKGLGDMFRFGKRRKDDK 1034
Db 807 ROMNGNOERKGDTRKRRKKTGKKKKDRKKKKKKGLGDMFRFGKRRKDDK 865

```

```

RESULT 9
Q9HC48 PRELIMINARY; PRT; 667 AA.
AC Q9HC48;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE CCLL TUMOR ANTIGEN SE2-5 (FRAGMENT).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=21143360; PubMed=11149944;
RA Eichmuller S., Usener D., Dummer R., Stein A., Thiel D.,
RT Schendendorf D.;
RT "Serological detection of cutaneous T-cell lymphoma-associated
RT antigens.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).
DR EMBL; AF177228; AAG33676.1; -.
DR HSSP; O12923; 3PZ;
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
FT NON_TER 1
FT TER 667
SQ SEQUENCE 667 AA; 73499 MW; C63EC16802BAE02 CRC64;

```

```

Query Match 47.4%; Score 3335; DB 4; Length 667;
Best Local Similarity 94.3%; Pred. No. 4.7e-181;
Matches 657; Conservative 3; Mismatches 7; Indels 30; Gaps 1;

```

```

QY 313 EHEULFRENOCIVRINCGLRNRFRFOAOHMRFOAMRTPILMFHVVPANKEOYDOLSDS 372
Db 1 EHEULFRENOCIVRINCGLRNRFRFOAOHMRFOAMRTPILMFHVVPANKEOYDOLSDS 60
QY 373 EKNYYSRSRPSDOYIDNRSVSAGLHTVORAPRLNHPEDQIDSHSLPHSAHPSGKRP 432
Db 61 EKNYYSRSRPSDOYIDNRSVSAGLHTVORAPRLNHPEDQIDSHSLPHSAHPSGKRP 120
QY 433 SAPASAPONEVSTVSSGNTKKGRLNIOLKGTGEGITSRDVYIGSAPVYKN 492
Db 121 SAPASAPONEVSTVSSGNTKKGRLNIOLKGTGEGITSRDVYIGSAPVYKN 180
QY 493 ILPRGAIIODGRILKAGDRILEVNGYDLGKSGEEVYSLRSTKKEGVYSLVFRQEDAFH 552

```

```

Db 181 ILPRGAIIODGRILKAGDRILEVNGYDLGKSGEEVYSLRSTKKEGVYSLVFRQEDAFH 240
QY 553 PRELNAEPQOMQIPRETKADEEDIVLTPDGTREFLTFEVLPLNDGSGAGLGVYAGNSKE 612
Db 241 PRELNAEPQOMQIPRETKADEEDIVLTPDGTREFLTFEVLPLNDGSGAGLGVYAGNSKE 300
QY 613 NHADGIFPKSLIINGAASKDGRILVNDQILVANGESILGKTNDAMETLRRSKTEGKN 672
Db 301 NHADGIFPKSLIINGAASKDGRILVNDQILVANGESILGKTNDAMETLRRSKTEGKN 360
QY 673 RCMOLIVARRISKCNELKSPGSPPELPETALDRERRRISHSLYSGIBSDSPRN 732
Db 361 RCMOLIVARRISKCNELKSPGSPPELPETALDRERRRISHSLYSGIBSDSPRN 420
QY 733 AALSRIMGESGKYQLSPVNMPODDVYIIEDDRPLVPLPHLSDQSSSSSHDDVGFYADAGTW 792
Db 421 AALSRIMGESGKYQLSPVNMPODDVYIIEDDRPLVPLPHLSDQSSSSSHDDVGFYADAGTW 480
QY 793 GTWAKAAISDSADCSLSPVDVYLAFOREGFGROSMSEKRTQFSDASQDLFVTRKSKS 852
Db 481 GTWAKAAISDSADCSLSPVDVYLAFOREGFGRO----- 514
QY 853 MDLGIADETKLTNTVDDOKAGSPSRDVGPSLGLKSSLSLESLQTAVAEYTLNGDIPFHRPR 912
Db 515 ---LADETKLTNTVDDOKAGSPSRDVGPSLGLKSSLSLESLQTAVAEYTLNGDIPFHRPR 570
QY 913 PRIIRGNCNESFRAIDKSYDKPAVDDDDGEMTLEEDTEESSRSGRESVSTASQPSH 972
Db 571 PRIIRGNCNESFRAIDKSYDKPAVDDDDGEMTLEEDTEESSRSGRESVSTASQPSH 630
QY 973 SLPRONGNOERKGDTRKRRKKTGKKKKDRKKKKK 1009
Db 631 SLPRONGNOERKGDTRKRRKKTGKKKKDRKKKKK 667

```

```

RESULT 10
Q96NX7 PRELIMINARY; PRT; 1205 AA.
AC Q96NX7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PARTITIONING-DEFECTIVE 3-LIKE PROTEIN SPLICE VARIANT A.
GN PAR3L.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gao L., Macara I.G., Joberty G.;
RT "Multiple splice variants of Par3 and of a novel related gene, Par3L,
RT produce functionally different proteins.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF426250; AAL30664.1; -.
SQ SEQUENCE 1205 AA; 132494 MW; 26E6704CCDE8CDB8 CRC64;

```

```

Query Match 29.3%; Score 2061.5; DB 4; Length 1205;
Best Local Similarity 37.9%; Pred. No. 1.7e-108;
Matches 539; Conservative 191; Mismatches 377; Indels 317; Gaps 49;

```

```

QY 1 MKVTYFGRTGRVYVPCGDGHMKVFSLIDAVTRYKRAIAKPNWIOVHRLEDGGLD 60
Db 1 MKVTYFGRTGRVYVPCGDGHMKVFSLIDAVTRYKRAIAKPNWIOVHRLEDGGLD 60
QY 61 LDDIICDVADDDKRIIVAVPDODPRHGGDGSASSTGTQSPERFSGSELGNNVSAFOPYQ 120
Db 61 PDDVADVDEDKILIAFEEDEPLHKTESPSGNPADROSADAEFEVAA-QLAAFRP-- 117
QY 121 ATSELEVPVSVLRAMPPLHVRSSDPALIGLSTSVSDSNFSSESPSKNPTKSTAGFL 180
Db 118 LGGEIEVTPSALKIGTPILVRRSSDP----- 143

```



```
QY 181 KONTAGSPKTCRCKDENRSLPRTDSNWSNOFORONARSSLSASHPMVKMLEQCODE 240
Db 144 ---VGP-----ADTOPSASHP--GGOSLKLVPDS 170
QY 241 DGTEDNSRVEPVGHADTGLEHPNF--SLDDMVKLVFVNDGGPLGIHVVP--FSARGR 297
Db 171 TQNLDR--EVLNGVQTELLTSPRTKDTLSDMTREIVELSGGGPLGIHVVPFSSLSGR 227
QY 298 TLGLVLRLEKGGKAHEHNLFRENDCIVRINDGLRNRRFEOAQHMFQAMRTPIIMFHV 357
Db 228 ILGLFRIGIEDNSRSKRGELFHENECIYKINNVLDVDTFAOAOVFFQAMKSPSVLHV 287
QY 358 VPANKEOYEOLSOSEKNYYSRSPSQYIDNRSVNSAGLHTVORAPRLNHPPEQIDS 417
Db 288 LPQNRQYERK-SVIGSLNIFGNNDGVLTQVPPVHGKSGLKT---ANLTGTDSPETA 343
QY 418 HSRLPASHAPS-----GKPPSAPASAPQNVFSTVSSGYNTKKIGKRLNQLKKGTEGLGF 473
Db 344 SASLQONKSPRVPRLGKRPSSPSLSP-----LMGFSNNNAKIKIDLKKGPEGJGF 395
QY 474 SITSRDVTIGSAPITYVNIILPRGAIIODGRKAGDRLEIVNGVDLVGKSQEBVSVLRS 533
Db 396 TVVTRDSSIHGPRPIFYVNIILPRGAIIKDRLOQSDRIIEVNGRVDYRTGTOBELVAMLRS 455
QY 534 TRMEGTVALVPRODAFHPRLENAEPSQOMQIPKTKAEDEDIVLTPDGTREFTFEVPL 593
Db 456 TKOGETAALVIAROGHFLPRELKEPDCALSTETS-----EQLTFEFLP 501
QY 594 NDSGSGAGVGSVKGRSKENHADGIFVKSTINGGAASKDRGLRVNDOLIANNGSLIGK 653
Db 502 NDSGSGAGVGSVKGRSKSETGTDLGIFKSTIHGAARFDGLRVNDOLIANNGSLIGK 561
QY 654 TNOAMETLRSMSTEGNRCGMQILVARISKCMELKSPGSPPELPIETALDRERR 713
Db 562 SNHEAMETLRSMSEGNIRGMQILVAR-----PERM----- 596
QY 714 ISHSLSIGIEIDESPSSRNALSRIMGESGYQLSPYVNMPODVTIIEDRLVPLPHL 773
Db 597 -----EDPACGAFSKPCFEN--CONAVTTSRRNDNSI-----L 628
QY 774 SDQSSSSHD-DVGFVYADAGTMAKAISDS-----ADCSLSPDVVPLAFQRE 821
Db 639 HPLGCSPODOKGGLLPNDG--WAESEVPSPPTPHSALGLGLEEDYSHSGSAVAYFPQ 687
QY 822 GFGROSMEKRTKOPSDASOLDVFVTKRSKSMGLGIDMETKLNTVDDQKAGSPSRDVGPS 881
Db 688 HINFRSVTPAR-----QPEISINLKASKSMDL--VPDESKVHSLAGKSESPSKDFGT 738
QY 882 LGLKSSSLESIQTAVAEVTLNGDIPEFHRPRITIRGRCNESPRAAIDKSYDKPAVDD 941
Db 739 LGLKSSSLESIQTAVAEVRKN-DLPFHRPRHMYVRGRCNESPRAAIDKSYDGP----- 792
QY 942 DEGMTELED--TESSSRRESVSTASDQPSHLSERQNGQEGDKT---DRKKK-- 994
Db 793 -----EELPADGISDSSSHSGGALNCESAPGNSLEDEMENKARVYKTKKEKKEK 848
QY 995 -TGKRRKDRKEDKMAKAGMLGCLDMFRFGKHKRD--DKIEKTKIKIOESFTSEE 1051
Db 849 LKVKKKKKRENEDEPERIKK--KGFQMLRFGKKKEDKGGKAQOKGLK--HGLRRE 903
QY 1052 EIRIKKOEOIRQAKTREREQARE-RDYA-----ETIODEHRTGCDDEL--MYGVSS 1103
Db 904 ELEKKKEERERIGAHQOELREKQARGLLDYATGALISYVDMO-----DEMPNYARVNH 958
QY 1104 Y-EGSMAUN-----ARPOS-----PREGH-----MDALAYOKKPNKSPVDSNR 1145
Db 959 FEEPCTSANVFSPSPRPFGPGYPRDGHPLSPERDHLGLTAKVKNKPYHPL--VPADSGR 1017
QY 1146 SPSNHDRIQRLROEFOAQOD-----EDVEDRRTYSPFQPMAPAPATQSGRHSVYE 1200
Db 1018 PGGSGTDRIOKARKEYQAARREGPFLVEDDEGRARSEYDLDLWVGR-GRDQNAHLARE 1076
```

```
QY 1201 VOMOROEERSSQOARQYSSLPROSARKNASVSQDSWEQNTSPG--EGFQSAKENR 1258
Db 1077 -----GMEQYASLPRGPAD-----PVDYLPAPRGLYKKEBELPY 1112
QY 1259 YSSYGSRNGYLGGHFNARVWLETOELLROQRKEDQMKQPSSEGSNTDYKKYVD 1318
Db 1113 YP-----GAHPMH-----PPKGSYPRPTELRVAD 1136
QY 1319 ---PSYAPP-----KGFRODVPSPSQVARNRLQTPPEKGRP 1353
Db 1137 LRPQHYPPPPAPQHKGPRRODVPSPQHQRMAYQ--ETGRP 1178
RESULT 11
Q96NX6
AC 096NX6; PRELIMINARY; PRT; 1143 AA.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PARTITIONING-DEFECTIVE 3-LIKE PROTEIN SPLICER VARIANT B.
GN PARL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Gao L., Macara I.G., Joberty G.;
RT "Multiple splice variants of par3 and of a novel related gene, par3l,
RT produce functionally different proteins.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF428251; AAL30665.1; -.
SQ SEQUENCE 1143 AA; 126102 MW; 368B69CF81D45E7E CRC64;
Query Match 25.9%; Score 1825.5; DB 4; Length 1143;
Best Local Similarity 35.0%; Pred. No. 3.8e-95;
Matches 498; Conservative 184; Mismatches 363; Indels 379; Gaps 49;
QY 1 MKVTYCFGRFVRVPCGDMKVFSLIOQAVRYKAIKDPNWTQVRLRHGCGILD 60
Db 1 MKVTYCFGRGTIVVPCKEQOLNVGELTQALQRYLTKREKGGYWKIHHLEYTDGILD 60
QY 61 LDDIICDVADDKRLVAVDEODPHHGCGDTASSTGTQSPETIFGSELGTNNVSAFYQY 120
Db 61 PDDVLADVVEDDKLVAVEDEQEPKHKIESPSGNPADRSPDAFEYVAA--QLAAKP-- 117
QY 121 ATSEIEVTPSVLRANPCLHVRSSDPALIGLSTVSDSNFSSEEPSRKNPTRMSTTAGFL 180
Db 118 ICGEIEVTPSALKLGTPLLYRRSSDP----- 143
QY 181 KONTAGSPKTCRCKDENRSLPRTDSNWSNOFORONARSSLSASHPMVKMLEQCODE 240
Db 144 ---VGP-----ADTOPSASHP--GGOSLKLVPDS 170
QY 241 DGTEDNSRVEPVGHADTGLEHPNF--SLDDMVKLVFVNDGGPLGIHVVP--FSARGR 297
Db 171 TQNLDR--EVLNGVQTELLTSPRTKDTLSDMTREIVELSGGGPLGIHVVPFSSLSGR 227
QY 298 TLGLVLRLEKGGKAHEHNLFRENDCIVRINDGLRNRRFEOAQHMFQAMRTPIIMFHV 357
Db 228 ILGLFRIGIEDNSRSKRGELFHENECIYKINNVLDVDTFAOAOVFFQAMKSPSVLHV 287
QY 358 VPANKEOYEOLSOSEKNYYSRSPSQYIDNRSVNSAGLHTVORAPRLNHPPEQIDS 417
Db 288 LPQNRQYERK-SVIGSLNIFGNNDGVLTQVPPVHGKSGLKT---ANLTGTDSPETA 343
QY 418 HSRLPASHAPS-----GKPPSAPASAPQNVFSTVSSGYNTKKIGKRLNQLKKGTEGLGF 473
Db 344 SASLQONKSPRVPRLGKRPSSPSLSP-----LMGFSNNNAKIKIDLKKGPEGJGF 395
QY 474 SITSRDVTIGSAPITYVNIILPRGAIIODGRKAGDRLEIVNGVDLVGKSQEBVSVLRS 533
```

```

Db      396 TVVTRUSSINGPPIPVKNLIPKGAIAIKGRIQSGDRILEVGRVOTGTOBEIYAMLR 455
      534 TKMGCTVSLVFERODFAHFRREINAEPSQMIPKTKADEIDIVLTGTRRFLFEVPL 593
      456 TKGCTVSLVIAIAREGHFLREL----- 478
      594 NDGSGAGLVGKGNRKENHADLGIIVKSIINGCAASKDGRILRVNDOLIAVNGSLIG 653
      479 -----DGRIRMDQOLIAVNGSLIG 499
      654 TTNODAMETLRSMSTEGNKGIMQILVARIKCNELKSPGSPPELPLEVALDRERR 713
      500 SNHEMELRLRSMWEGNIRGMIDLVILRR-----PERRP----- 534
      714 IHSLSLSTEGIDESPNNALSRIMESGKYQLSPVNNPDDTVIIEDRLPVLPHL 773
      535 -----EDPAECGAFSKPCPEN--CQNAVTTSRNDNSI-----L 566
      774 SDQSSSSHD-DVGFVTADAGTAKAALSDS-----ACCSLSPVDVPLAFORE 821
      567 HPLGTGSPQDKOKGILLIPMDG-WAESEVPSPPTPHSALIGLEEDYSHSSGVDAAVYEPDQ 625
      822 GFGROSMSEKRTKOFSDASOLDVFTVRKSKSMDLGIADETKLNTVDOKAGSPSRDYGPS 881
      626 HINFRVTPAR-----QPSINLKASKSMDL--VPDESKVHSLAGCKSPSPKDEGPT 676
      882 LGLKSSSLESLQTAVAEYTLNGDIPFHRPRRIIRGCGNCSFRAALDKSYDPAYDD 941
      677 LGLKSSSLESLQTAVAEYRKN-DLPHFRPRHVMNRCNCSFRAALDKSYDPG----- 730
      942 DEGMETLEED--TESSRSRGFSVSTADQPSHSLEROMNGOEKDKT--DRKKK-- 994
      731 -----EELFAGLSKSSSHSOGALNCESAPQNSLEDENKARKYKKEKKEK 786
      995 -TGKCKKRDREKDKMKKAKGMLKGLDMFRGKHKRD--DKIEKTGKIIOESFTSEE 1051
      787 LKVKKKKKKENEDPERIKK--KGFAMLRGKCKKEDGKAEBQKTLK--HGLRRE 841
      1052 EIRIRKQOEIRLOAKTRFREFROARE-RDYA-----EIODHRTGCDDEL--MGCVSS 1103
      842 ELEKKKEERIRGAKHOLREKQARGLDYATGALISVYDM-----DDMDPNVARKH 896
      1104 Y-EGSMALN-----ARPOS-----PREGH-----MMDLVYAOVKKPRNSKSPSYDSNR 1145
      897 FHEPCTSANVFRSPRPAGRGPGYPRDGHPLSPERDHLGLIYAKVNNRPHPL--VPADSGR 955
      1146 STPSNDRTORLROEFOAKOD-----EDVEDRRRTYSFEOPFNPAPRATQSGHVSVE 1200
      956 PTGGSFDRLOKRLKEYQARRRGFPPLYEDDEGRARSPSEYDILWVGR--GPDGNANHLNFE 1014
      1201 VOMQROROBERESSQOAROYSSLPQOSKKNASSVSQDSWEONSPG--EGFQOSAKENPR 1258
      1015 -----GMEQVAYSLPRGPA-----PYDYLPAAPRGILYKERELPY 1050
      1259 YSSVQSGRNGYLGCHGFNAVMLETOELLROBORRKEQOMKOPRSEPSNYSYKKYOD 1318
      1051 YP-----GHNPRH-----PPKSYPRPTELKAVAD 1074
      1319 ---PSYAP-----KGFPODYPPSPSOVAYRLNRLQTEPKGRP 1353
      1075 LRYOHYVPPRPAOHKGRFRODVPSPPOHAPAYQ--ETGPR 1116

```

```

RESULT 12
O9NM14 PRELIMINARY; PRT; 347 AA.
AC 09NM14;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA FLJ20754 FIS, CLONE HEP02246 (UNKNOWN) (PROTEIN FOR
MGC:195181).

```

```

OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_Taxid=9606;
RN 11
RP SEQUENCE FROM N.A.
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Odayashi M.,
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK000761; BAA91366.1;
DR EMBL; BC011711; AAH1711.1;
SQ SEQUENCE 347 AA; 40538 MW; BE2B3557996EC91E CRC64;

```

```

Query Match 25.9%; Score 1822; DB 4; Length 347;
Best local similarity 100.0%; Pred. No. 1.2e-95;
Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      1010 MKAKKGMKLGDMFRGCKHKKDKIEKTKIKIOESFTSEERIRMKOBERIOAKTRE 1069
      1 MKAKKGMKLGDMFRGCKHKKDKIEKTKIKIOESFTSEERIRMKOBERIOAKTRE 60
      1070 FRRQARERDYAEIODFHRFTGCDDELMTYGVSSYBSGSMALNARPOSPPRGHMDALYNO 1129
      61 FRRQARERDYAEIODFHRFTGCDDELMTYGVSSYBSGSMALNARPOSPPRGHMDALYNO 120
      1130 VKKPRNSKSPVDSNRSTPSNDRTORLROEFOAKODEVEDERRRTYSFEOPWPNARPA 1189
      121 VKKPRNSKSPVDSNRSTPSNDRTORLROEFOAKODEVEDERRRTYSFEOPWPNARPA 180
      1190 TOSGRHSVSEVOMQROROBERESSQOAROYSSLPQOSKKNASSVSQDSWEONSPGEG 1249
      181 TOSGRHSVSEVOMQROROBERESSQOAROYSSLPQOSKKNASSVSQDSWEONSPGEG 240
      1250 FOSAKENPRYSYVQSGRNGYLGCHGFNAVMLETOELLROBORRKEQOMKOPRSEGPSN 1309
      241 FOSAKENPRYSYVQSGRNGYLGCHGFNAVMLETOELLROBORRKEQOMKOPRSEGPSN 300
      1310 YDSYKTKYODPSYAPPKGFRQDVPSPSOVAYRLNRLQTEPKGRPYS 1356
      301 YDSYKTKYODPSYAPPKGFRQDVPSPSOVAYRLNRLQTEPKGRPYS 347

```

```

RESULT 13
O96DK9 PRELIMINARY; PRT; 624 AA.
AC 096DK9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA FLJ25236 FIS, CLONE STM02096 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_Taxid=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=GASTRIC MUCOSA;
RA Ninomiya K., Wagatsuna M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furiya T., Takahashi M., Kikawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Suzuki Y., Hata H.,
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA Nagai K., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.

```



```

Db 225 SSGNHOPFASGRISMOFLDNGCYKMEAEKLNQPAQOYQOQSHHAGHONGAYS 284
OY 246 -----DNRSEVYGHADTGLEHIFPESLDDMKVLEVNDGPGIINHVPVSAGCRTL 299
Db 285 SKSLPRESKREPELGAQYESIRE-----KDGEMLLINIEYSGPLGLTALPKKEHG--- 335
OY 300 GILYKRLKLEGGKAHEHLEFNDCIVRINDGDELRNRFEOQAMERQAMRTPIIHFHYVP 359
Db 336 GILYOHVERGSRRAEGRRL-RDDDRILINIGIKLIGLESQVQOLRRALLE----- 384
OY 360 AANKQOYBOLSQSEKNYYSRSFSDSOYIDNKNVSNAGLTHQVAPRLNHPPEQIDSHS 419
Db 385 -----SSELRYRVLKGRNR-----RQORDSKVAENEVAIYS----- 417
OY 420 RLPHSAHPSGRPPAPAPAPQNVFSTVSSGYNTKIKIGKRLNIOJLKKTEGLGFSITSRD 479
Db 418 -----PTRKPHAAPVGTSLQY-----AMTRKIGRKLEIMLKKGPNGCLGFSVTRD 462
OY 480 VTIIGSAPTYVKNILPRGAIDQRLKAGDRLIEVNGVDLVGKSOEEVVSILRSTKMEGT 539
Db 463 NPAGGHCPYIKNILPRGAIEDGRILKPDRLLEVDTGTPMGTQYTDVVALRGMAGAT 522
OY 540 VSLIVFRQ-----BDAPHPRELNA-----EPS----- 561
Db 523 VRIYVSRQOQLAEQADQAPKSGAVAPVAPVAPAAAAPPPIPVQKSSASRLFTH 582
OY 562 --QMOIQR-----ETKAEDEDIVLPDG--PREPITFEVPLINDSGSAGIAGVYK- 607
Db 583 QOQOQLNESOHFI DAGSENASNDSPSPSNMWSHREELTHIPVHDIEKALIGVSYK 642
OY 608 -----NRKENHADIGIFVYSIINGGAASKDRLRVNDQILYVNGESIL 651
Db 643 TCSNLNASSGASSGSLKMHODDGLIEFKVNIHGAASRQGRILMNDQLLSVNGVSLR 702
OY 652 GKTNDAMETLRSM-STEGNKRMIOIYARRISKC----- 687
Db 703 GONNAEAMETLRAMVNTPGKHPTITLLVGRKILKRASSDILLDSHSHSSNSGG 762
OY 688 -----NELKSPGSPGELPITFALDRERIRISHLSYSGIEGLDPSRNALSR 738
Db 763 SNSNSGNNNSSNASNSGATYIYLSPEKREORCNGGGGAGENMRMSNPVLDL 822
OY 739 MG-----ESGKYQLSPVYNNPO----- 755
Db 823 TGGICSSNSAOPSSQOOSHQOOPHSQOQOORLPAAPVCSAALRNEYYMATNDWSP 882
OY 756 -----DDVILIEDRLVLP-----PILSDQSSSSH-----DYGfV 788
Db 883 AQNHILMTAGNTALLIEDDAEPMSPTLPPARPHDGCHCTSSANPSQNLAVNGOPPIINTV 942
OY 789 TADAGTWAKAALSDADCLSPDVDPVLAFOREGFGROSMSRK-----RTK 834
Db 943 PGTPTSTNSNDATYSSQSLLETN-SCVEHFPSDALGRNISISKHHAALDARETGYONK 1001
OY 835 QFSDASOLD-FVKTRKS-----KSMGLGLAD-----ETKLN-T 865
Db 1002 KLREPERERRIQUTKSAYVGSIESITARIASANAQFSGYKNAKTASITEOREQOOLA 1061
OY 866 VDDOKAGSPRDVPSLGKSSLSLSTAVAEVTINCDIPFHRPRLIRGRCNBSF 925
Db 1062 AAEEARDOJLDLIGPLGKMKSSLSLSTQMYELOMS-DEPRHQALRAPRGREGESL 1120
OY 926 RAADIKSYD--K--AVDDDDGMEETLEEDTEESSRSRGRESVSTRASQPSHLSERON 979
Db 1121 RAAVSEPDASKPRKTMLEDCD-----HEGFAASORNG-----PROSSIN 1161
OY 980 GNOEGDTRDKKDKTKGKKDKDKKAKKMLKGLGDMFRGKHKDKDKLEKTKG 1039
Db 1162 DGH-KCKSSRAK-----KPSILRIGCHMFRGKHKRGKGVAVPVN 1200
OY 1040 ---KIKIOSFTSEERIRIMQOEDERIOAKTREFREROQAREBYAELQDHRFGODEL 1096
Db 1201 YAVNISPTSYSTATSPQLQOQO-----QOLOHOQO----- 1235

```

```

OY 1097 MYGVSVSYEGSMALNARQSPREGHMDALYAQYKPKRNSKSPVDSNRSTPSN----- 1150
Db 1236 -----OIPYAA--LAALERNGKPRYOPPPPLPAPNGVSGIHOND 1275
OY 1151 -----HDIORLROEF--QOAKODEVEDERRRRTTYFEOPMPNARP 1188
Db 1276 IFNHRYOHTANTYEDLHQOHOQOHSRHOHYHSGRSARSQDV-SMHSYSSGSGSIAOP 1334
OY 1189 ATOS-GRHSYS-----VEVOMORORERESSQOARQYSSLPROSRKNASSVSODSWE- 1241
Db 1335 QAQSNQVPRMPSYEEYETVOOQRVGSIKSHSSA-----TSSSSPINVPHKA 1384
OY 1242 ---QNTSPGEGFQSAKENPRYSYSGSRNGYLGHGFNARVMLETQELLROE----- 1290
Db 1385 AAMNGYRSPASINSAR-----SRGPFV-----TOYVITREGSGGIPAH 1422
OY 1291 --QRRKEQOMKROPS-----EGSNTDS 1312
Db 1423 LLQHQOQOOLQOQOQPTQYQYQKMSGPSQYGS 1453

```

```

RESULT 15
ID 096782 PRELIMINARY: PRT: 1464 AA.
AC 096782:
DT 01-MAY-1999 (TREMblrel, 10, Created)
DT 01-MAY-1999 (TREMblrel, 10, Last sequence update)
DT 01-JUN-2001 (TREMblrel, 17, Last annotation update)
DE BAZOOKA PROTEIN.
GN BAZ OR BAZOOKA OR CG5055.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuchinke U., Grave F., Knust E.;
RT "Control of spindle orientation in Drosophila by the Par-3-related
RT p21-domain protein Bazooka."
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ130871; CAI10224.1; -.
DR HSSP: Q12923; 3PDZ.
DR Flybase: FBgn000163; baz.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 3.
DR SMART: SM00228; PDZ; 3.
DR PROSITE: PS50106; PDZ; 3.
SQ
SEQUENCE 1464 AA; 157398 MW; 77381A5FBCA4C4B CRC64;

```

```

Query Match 16.38; Score 1149.5; DB 5; Length 1464;
Best Local Similarity 26.28; Pred. No. 11e-56;
Matches 432; Conservative 212; Mismatches 470; Indels 537; Gaps 61;

OY 1 MKYTVCFGTRVYVCGGCHMKVFSLIQAVTRYRKALADPNYWIQVHLEHGDDGILD 60
Db 1 MKYTVCFGTRVILPCGSEELVRLVKEATRRYIKAGK-PDSWVTVTHQ-TQSGILD 58
OY 61 LDDILCDVADDKDRILAVFDEOD-----PHNGGOST-ASSTGOSPEIFGSELGTNNVS 114
Db 59 PDCQVADVADDRQILAHDPDDPGDPGVQGGDASSSSVGSGSPDIFRDP--TTEA 116
OY 115 AFQPIQ-ATSEIVTPS-----VLRAVMPLVHRRSDPALIGLTSVSDSNFSSEPSRK 169
Db 117 PTCPRDLSTPHIEVITSTSPNAGLVGLMVRSSDPNLLA-SLKAEGSN----- 165
OY 170 PTRWSTAGF-----LKONTAGSPKTCDDKDN-----YRSLPRDTSNMS 210
Db 166 -KNSAAPHAGDSPERFLFDKAGQOLSPQWEEEDDSDSHQKROLHQOPPHANGS 224
OY 211 NOFORDNARSSLASHMVG-----KMLEKQ--QDEDETEE----- 245

```

```

Db 225 SGNHOPPARSRSLMOTIGDGNNGYKMMBAEKLDONPPAQOTYQCGSHHAGONGAYS 284
QY 246 -----DNSRVEPVGHADTGLEHIFNPSLIDMWKLYEVNDGPGPLGIHVPPSARGRTL 299
Db 285 SKSLPRESKRKPELQAYVSIRES-----KQDEMILLIINVEYSPGLGTLPLPKERHCG----- 335
QY 300 GLLVRLKLEGGAEHBNLFRENDICVIRINDGULRNRRFQAOHMRQAMRTPIIHFVHPV 359
Db 336 GLLVHVEGSAERQOL-RDDRIELINGIKLIGTESQVOEOLRALE----- 384
QY 360 AAMKOEYEOLOSSEKNNYSSRFPDSOYIDNRVSASGLHUYQARPLNHPPEQIDSMS 419
Db 385 ---SSELRAVVLGDRNR-----ROQRDSKVAQMEVATVS----- 417
QY 420 RLPHSAHSGKPPASAPAPQWVFSTTVSSGYNTKIKGRKLNIOLKKEGLEFSTSD 479
Db 418 -----PFRKPHAAPVGSLOV-----ANTRLKLGKIQIMLKKGNGLGFSVTTMD 462
QY 480 VTIGSAPLYVKNILPRGAIODGLKACDRILEVNGVDLVKSGQREVVSLRSTMEGT 539
Db 463 NPAGAHCPYIKNIILPRGAIEDGRKPGDRILEVDGTPMTGKTQDVVAIILRGMPAGAT 522
QY 540 VSLVFRQ---EDAFHPELMA---EPS----- 561
Db 523 VAIIVSRQOELAEQADQPAEKSAAGVAVAPPAVPAAPAPPIPVOKSSANSRLFTN 582
QY 562 ---QMOIPK-----ETKAEDEDIVLPDG--TREFLTFEVLNDSGAGLGVSVKG- 607
Db 583 QOQSOLNESQHTIDAGSEEAASNDLPPSSNSMHSREELTLHPVHDTERKAGLGVSVKK 642
QY 608 -----NRSKENHADLGIFVKSIINGGAASKDRLRVNDQILAVNGESLL 651
Db 643 TCSNLNASGSSASGSMGLMKHDGDLGIFVKVHIGGAASRDGRLRMNDQLSVNGVSLR 702
QY 652 GKTNDAMETLRRSM-STEGNKRGMLOLVARRISC----- 687
Db 703 GONNAEMETLRRAVNTFGKHPGTTLLVGRKILRSASSDILDSHSHSHSSNSGG 762
QY 688 -----NELKSPGSPPELPIETALDDRERISHLSYIGIEGLDESPRNALSR 738
Db 763 SMSNGSGNNSSNASDMSGATVYILSPERKREQKONGGGGSGACNEMNRMSNPVLDRL 822
QY 739 MG-----ESGKYQLSPVNMPO----- 755
Db 823 TEGICSSNASQPSOQSHQOPHPSOQOQORRLPAAPVCSAALRNESYMATNDNMP 882
QY 756 -----DDTVIIEDDRPLVLP-----PHLSDOSSSSSHD-----DVGTV 788
Db 883 AOMHLMTAHGNLTALLIEDAEPMSPTLPARPHDGOHCNTSSANPSONLAVGNOGPPINTV 942
QY 789 TADACTWAKAALSDADCSLSPDVPVLAFOREGFGROMSK-----RTK 834
Db 943 PCTPSTSSNFDATYSSQLSLETN-SGVEHFSRDLGRSISERKHNAALDARETGITYQRNK 1001
QY 835 QESDASQLD-EVYTKRS-----KSMDLGLAD-----ETKLN-T 865
Db 1002 KLREERERERRIQLTKRSAYVGGISILFARIASANAQFGYKHAKTASIEQRETQOLA 1061
QY 866 VDDKAGSPRDVPSLGKSSSLESLOTAYAEVTLNDIDPFHRPRPIIRGRCNESP 925
Db 1062 AAEEARQDGLGSPSLGKKSSSLESLOTWQVLOLMS--DEPRHQALRAPRGREDSL 1120
QY 926 RAAIDKSYD--KP-----AVDDDEGEMTLEEDTESSRSGRSVSTASQPSHSLEROM 979
Db 1121 RAAVSEPDASKPRKTWLEDD-----HEGFGASQRNG-----PFQSSLN 1161
QY 980 GNOEKGDTDRKKDKTKGKKKDRKDKKAKKMLGLGDMFEGKHRRDKLIEKTG 1039
Db 1162 DGH--GCKSSRAK-----KPSILRGIGHMFRFGKRRKDGVPVDN 1200
QY 1040 ---KIKIESFTSEERIRMKOEORLQAKTRERERQARERDYAEIQDFHRTFGCDDDEL 1096

```

```

Db 1201 YAVNISPTSVSTATSPOLQOQOQ-----QOLQOHQOQO----- 1235
QY 1097 MYGVSYSEGSALNARPOSFREGHMDALYAQVKPRNSKPSVPDSNRSTPSN----- 1150
Db 1236 -----DIPTA--LAALERNGKPRVQPPPLDAPRNKNGVSGNHIQND 1275
QY 1151 -----HDIRIOLRQEF--QAKKODEVEDRRRTYSFEQPPWPNAP 1188
Db 1276 IFNHRYQHYANEDLHQOHQOHQISRRHQHYHSQARSQDV--SMHSTSSGQPGSLAQP 1334
QY 1189 ATQS--GRHSYS-----VEYOMORQROBERESSQQAQROYSSLPRSKNASSVSQDSWE- 1241
Db 1335 QAOSDGVPRMSSYIETVQOQVGSIKSHSSA-----TSSSSPTNVPMWKA 1384
QY 1242 ---QNTSPGEFQSAKENPRYSSYOGSRNGYLGHGFNARVMELETOELLROE----- 1290
Db 1385 AAMNGYSPASLNSAR-----SRGPFV-----TQVTTREQSSGGIPAH 1422
QY 1291 ---ORRKEQOKKOPPS-----EGPSNIDS 1312
Db 1423 LQOHQOQOQOQOQPTQYQVOKMKSGPSQYGS 1453

```

Search completed: July 24, 2002, 14:33:22  
 Job time: 253 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:19:24 ; Search time 45.21 Seconds  
(without alignments)  
2882.042 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 7035  
Sequence: 1 MKVTVCFGKTRRVVPCGDGH.....SQVARNLRLOTPEKGRPFYS 1356

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6332.5	90.0	1337	2 T13948	atypical protein k
2	1149.5	16.3	1464	2 T13716	baculovirus gene prote
3	547	7.8	1360	2 T34302	cell polarity prot
4	361.5	5.1	2054	2 T46612	multiple PDZ domain p
5	352.5	5.0	2055	2 T30259	multiple PDZ domain p
6	335.5	4.8	2172	2 T20145	protein tyrosine p
7	332	4.7	2466	2 T67629	protein tyrosine p
8	328.5	4.7	2450	2 T51625	protein-tyrosine-p
9	323	4.6	1131	2 T15617	hypothetical prote
10	321	4.6	2490	1 A54971	protein-tyrosine-p
11	314	4.5	2294	2 T67630	protein tyrosine p
12	296.5	4.2	5327	2 T13564	microtubule-associ
13	295	4.2	926	2 T38756	homolog of Drosoph
14	293	4.2	767	2 T09599	postsynaptic densi
15	293	4.2	1095	2 T43275	neurabin - rat
16	289.5	4.1	1410	2 T47137	hypothetical prote
17	286.5	4.1	904	2 T38757	homolog of Drosoph
18	286	4.1	724	2 JH0800	postsynaptic densi
19	286	4.1	870	2 G01974	channel associated
20	283.5	4.0	628	2 T09458	numb-binding prote
21	283.5	4.0	728	2 T09457	numb-binding prote
22	279	4.0	852	2 T10811	channel associated
23	278.5	4.0	720	2 A45436	synapse-associated
24	278.5	4.0	911	2 T56552	synapse-associated
25	278	4.0	1256	2 JEO209	brain-specific ang
26	272	3.9	1171	2 T42372	probable guanylate
27	260	3.7	1829	2 T41751	1-afadin - rat
28	250	3.6	1012	2 T23160	hypothetical prote
29	246	3.5	1663	2 T42092	s-afadin - rat

30	245.5	3.5	1553	2 T03301	rab3 effector prot
31	241.5	3.4	3488	2 T34418	hypothetical prote
32	240.5	3.4	1277	2 T14152	synaptic scaffold
33	239.5	3.4	2248	2 A35938	profilaggrin - hum
34	237.5	3.4	3924	2 S37431	ankyrin 2, neurona
35	228.5	3.2	1112	2 T32733	AMPA glutamate rec
36	228	3.2	2464	1 QRMSP1	microtubule-associ
37	227.5	3.2	960	1 A39651	discs-large tumor
38	226	3.2	2526	2 T20531	hypothetical prote
39	226	3.2	2722	2 T20532	hypothetical prote
40	226	3.2	2738	2 E88320	protein F07A11.6 f
41	226	3.2	6642	2 T29757	protein UNC-89 - C
42	225	3.2	754	1 JC5314	CDG28/cdg2-like ki
43	221.5	3.2	1281	2 T00346	hypothetical prote
44	221.5	3.1	771	1 A33430	h-caldesmon - chic
45	221.5	3.1	1666	2 T43169	hypothetical prote

ALIGNMENTS

RESULT 1	
T13948	atypical protein kinase C isotype-specific interacting protein ASIP - rat
C:Species: Rattus norvegicus (Norway rat)	
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000	
C:Accession: T13948	
R:Zumli, Y.; Hirose, T.; Tamai, Y.; Hirai, S.; Nagashima, Y.; Fujimoto, T.; Tabuse, Y.	
J:Cell Biol. 143, 95-106, 1998	
A:Title: An atypical PKC directly associates and colocalizes at the epithelial tight	
A:Reference number: Z17827; MUID:96437350	
A:Accession: T13948	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-1337 <ITD>	
A:Cross-references: EMBL:AB005549; NID:93868777; PIDN:BA43216.1; PID:93868778	
A:Genes: asbp	
Query Match	90.0%; Score 6332.5; DB 2; Length 1337;
Best Local Similarity	89.7%; Pred. No. 7e-309;
Matches 1216; Conservative 63; Mismatches 58; Indels 19; Gaps 2;	
QY 1 MKVTVCFGKTRRVVPCGDGHMKVFSLIQAVTRKKAIAKDPNMYIQVRLHEHGDGILD 60	
DB 1 MKVTVCFGKTRRVVPCGDGRMKVFSLIQAVTRKKAIAKDPNMYIQVRLHEHGDGILD 60	
QY 61 LDDIICDVADDKRLVAVFDEDDPHHGGDGTSAASSTGTQSPETFGSELGTNNVSAFQPYQ 120	
DB 61 LDDIICDVADDKRLVAVFDEDDPHHGGDGTSAASSTGTQSPETFGSELGTNNVSAFQPYQ 120	
QY 121 ATSEIEVPSVIRANMPILHVRSSDPALIGLSTVSQDSNFSESEPRKPTWSTAGTL 180	
DB 121 ATSEIEVPSVIRANMPILHVRSSDPALIGLSTVSQDSNFSESEPRKPTWSTAGTL 180	
QY 181 KONTAGSPKCDRKKDENTRSLPRDTSNMSNOFORDNARSSLSASHPMVGMKLEQODE 240	
DB 181 KONTAGSPKCDRKKDENTRSLPRDTSNMSNOFORDNARSSLSASHPMVGMKLEQODE 240	
QY 241 DCTEDNSRVEPVGADTGLEHHPNFSLDPMYKLVFVPNDGPGLGTHVVPFSAAGGRTLG 300	
DB 241 DCTEDNSRVEPVGADTGLEHHPNFSLDPMYKLVFVPNDGPGLGTHVVPFSAAGGRTLG 300	
QY 301 LTVKRLKGGKAEHNLFPRENDICVIRINDGLRNRFEDQAHMFQAMRTPLIMFHVVA 360	
DB 301 LTVKRLKGGKAEHNLFPRENDICVIRINDGLRNRFEDQAHMFQAMRTPLIMFHVVA 360	
QY 361 ANKEQYEQLSQSEKNNYSSRPSDQYIDNRSSVAGLHTVQARPLRNHPPEQIDSHSR 420	
DB 361 ANKEQYEQLSQSEKNNYSSRPSDQYIDNRSSVAGLHTVQARPLRNHPPEQIDSHSR 420	
QY 421 LPHSAHPGSKRPPASAPASAPQNVFTTVSSGYNTKKIGKRLNIQLKGTGGLFSTISRQV 480	
DB 421 LPHSAHPGSKRPPASAPASAPQNVFTTVSSGYNTKKIGKRLNIQLKGTGGLFSTISRQV 480	

A:Reference number: Z17708  
A:Accession: T13716  
A:Status: preliminary; translated from GR/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1464 <KUC>  
A:Cross-references: EMBL:AJ130871; NID:e1363519; P1D:e1363520; PIDN:CAAI0224.1  
C:Genetics:  
A:Gene: bazooka  
A:Cross-references: FlyBase:FBgn0000163  
A:Map position: X

```

Query Match 16.3%; Score 1149.5; Un-
tered Local Similarity 26.2%; Pred. No. 1.3e-49;
Matches 432; Conservative 212; Mismatches 470; Indels 537; Gaps
61

QY 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 60
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 61 LDDITCDVAADDEKRLVAFVFEQD-----PHHGDSGT-SASGTGSPSEFGSELTNNVS 114
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 59 PDDCVRDAVADDEQQLAFHFDPGDPGVGGGGAGSGSSVGTGSPDLFRDP--TNEEA 116
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 115 AEPQVQ-ATSEIETPS-----VLANNPLHRRSDPALGLSTVSDFNSFSEERSRKN 169
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 117 PTCPRDLSTPIHEVSTTSGPMAGLGVLGMKSSDPNLLA-SLAKESGN----- 165
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 170 PTWSTTAAE-----LKONTAGSEPTCDRKKDN-----YRSLPRTSINMS 210
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 166 -KRWSAAPHYACGDSPELFLIDKAGCGSLQPMWEEDDDPSHOLKEQLHQOPHAANGS 224
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 211 NQFORNARSSLSASHPMVG-----KLEKDE--QDEEGTGE----- 245
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 225 SSGHQDFPARGSLSMQFLGDNGYKWEAEKLNQOPPAQQTVOQSGSHAGGAGVVS 284
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 246 -----DNSRVEVYGAADTGLHITPNESLDDAVKLVEPNDGPGLGITHVVPFARGRL 299
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 285 SKSLPRSKRKEPLGQAYESIRE-----KQEMELLINETSGLTALPDKHEGG-- 335
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 300 GLVLYRLEKGGKAHENLRENDCTYRINDGLRRRRRQAOHMEFQOMRPTPIWFHVP 359
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 336 GLVQVHEPGSRARQGL--RRDRILEINGIKLIGLTESQVQEOURLALE----- 384
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 360 AANKEQVEQLSQSKENNYSSRFPDSQYIDNRSVNSAGLITVQAPRLNHPPEQIDSHS 419
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 385 --SSELRYKYLKGRNR-----RQQRDSKVADMEVATVS----- 417
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 420 RLPHSAMPSPKPSAPASAPQNVFTTSSGTYTKIKGRKLMDLKKGEIGESITSRD 479
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 418 -----PTRPKAAPVGTSLQV-----ATRRLGKRLQIMLKKPNGSGFVTTRO 462
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 480 VTIGGSAITVYKNTLPGALADQRLKAGDRLTEVNGVDLVGKSGQEVVSLIRTKMEGT 539
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 463 NPAGAHCPITVINKLIPGALIEDRLKPGRLLEVQGTPTMGKTQIDVYALIKGMPAGAT 522
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 540 VSLIVFRQ-----EDAFHRELNA-----EPS----- 561
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 523 VRLVYSRQQLAEQADQPAEKSAAGVAVAPAVPAAAPAPPIPVYKSSASSSLFTH 582
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 562 --OMQIPK-----ETKAEDEDIYLTTPG--TRELFLEFVPLNDSGSAGLGVSVK- 607
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 583 QQQSQNLNESQHFIDAGSEASNSDLPPSSMSHRSRELLHLHPVDTEKAGLGVSVKK 642
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 608 -----NRSKENHADLGIFVSIINGAGASKDGLRKYVNOQLIYVNESIL 651
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 643 TCSNLNAGSSGSSAGSGMLKHDGDLGIFVKNVHIGGASHDGRLRMNDQLLVNGVSLR 702
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 652 GKTNDQAMETLRSM-STEGNKRGMQIOLIVARISK----- 687
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 703 GONNAEMETLRAMVNTPGKHPGTTLLVGRKILRSASSDIIHDHNSHSHSSNGG 762
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58
QY 688 -----NELKSPSPQPELPIETALDRERRISHSLVSGIEGLDESPRNALSR 738
DB 1 MKVTGCEERTYVPPCGDGMKVFSLIQAVTETRYKAIAKDNPYIOWHRLHEHGDGILD 58

```



```

Db 763 SNSNGSGNNSSNASDNGATVYIYLSPEKRRQRCNGGGGAGNENMNSNPVLDRL 822
Qy 739 MC-----ESGKYOLSPYVNMPO-----755
Db 823 TGCCTSSNAGPSSQOOSHOOQPHPSOQOQORRLPAAPVCSAALRNESYMATDNMSP 882
Qy 756 -----DDVTIIEDRLPLP-----PHLSDQSSSSHD-----DVGCV 788
Db 883 AQMHMLTAHGNALLIEDDAPMSPFLPARPHDGCNTSSANPSQNLAVNGOPRINTV 942
Qy 789 TADAGTMAAALSDADCSLPDVPLAFORCGRGOSMSK-----PRK 834
Db 943 PGTPTSSNFDATYSQSLFTN-SCVEHPSDALGRSISSEKHNAALDARETGYQRMK 1001
Qy 835 QPSDASOLD-FYKTRKS-----KSMQDIAD-----ETKLN-T 865
Db 1002 KIREBERERRIQLTKSAVYGCSTESLTRIASANAOFQYHAKTASSIEORETQOOLA 1061
Qy 866 VDDOKAGSPSPDVPSLGLKSSSLESLOTAYAEVTLNGDIPFHRPRPIIRGCGNESF 925
Db 1062 AAEEARLDGLDGLPSLGLKSSSLESLOTAYAEVTLNGDIPFHRPRPIIRGCGNESF 1120
Qy 926 RAADIKSYD-KP-----AVDDDECMETLEEDTESSSGRESYSTASDQPSHSLEROMN 979
Db 1121 RAAVSEPPASKPRKRTWLEDGD-----HEGCFASQRMG-----PFGSSLN 1161
Qy 980 GNOEKDGTDRKKDKTKGKKKKDRKDKMKAKMKMLKGLDMFEGKHKKDKTEKGT 1039
Db 1162 DGRH-GCKSSRAK-----KPSILGIGHMFFGKNNKRDGVYVUN 1200
Qy 1040 ---KTIQESTSEERIRMKOEIRIQAQTRFRRQARENDYAIODFHRTFGCDEL 1096
Db 1201 YAVNISPTSPVSTATSPQLQOQOQO-----QQLQOHOQOQO-----1235
Qy 1097 MNGCVSYSGSALNARPOSPREGHMDALYAQVKKPRMSKSPVDSNSTPSN-----1150
Db 1236 -----OIPYAA--LAALERNGKPPAYOPPLPLPAPGVGSGINGIIOND 1275
Qy 1151 -----HRIQRLRQEF--QAQKODEVEDRRRTYSFEQPMFNARP 1188
Db 1276 IHNHRYQHANTEDLHQOHOQOHSRRHQHYSQARSQDV-SMSTSSGSPGSLADP 1334
Qy 1189 APOS-GRHSVS-----VEVOMORQORESSQOAOQOYSSLPGRKKNASVSQDSWE- 1241
Db 1335 QAQSDQVRPMSSYXEYEVQOQOQVGSIKHSHSSA-----TSSSSSPINVHMKA 1384
Qy 1242 ---QMSPEGQSAKENRYSYSGSRNGYLGCHGFNARVMLETDLRQF-----1290
Db 1385 AAMNGTSPASLNSAR-----SRGPFV-----TQVTIRESSGIGIPAH 1422
Qy 1291 ---QRRKEQOMKROPSS-----EGSPNYDS 1312
Db 1423 LLDQHOQOOLQOQOQPYQVQOKMSGRPOYGS 1453

```

```

RESULT 3
T34302
cell polarity protein par-3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34302
R:Bentley, D.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of C. elegans cosmid f54e7.
A:Reference number: 221502
A:Accession: T34302
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1360 <BEN>
A:Cross-references: EMBL:U00067; PIDN:AACT7513.1; GSPDB:GN00021; CESP:F54E7.3
A:Experimental source: strain Bristol N2; clone f54e7
C:Genetics:

```

```

A:Gene: par-3; CESP:F54E7.3
A:Map position: 3
A:Introns: 52/3; 102/1; 352/1; 411/1; 783/1; 832/1; 906/2; 990/3; 1049/2; 1265/2; 130
Query Match 7.8%; Score 547; DB 2; Length 1360;
Best Local Similarity 22.4%; Pred. No. 1.6e-19;
Matches 318; Conservative 170; Mismatches 444; Indels 490; Gaps 63;
Qy 46 IOVHRLR-EGDGGIIDLDDILCDVAD-DKDRLYAVNDE-----ODPH-----86
Db 22 IHHVRLCASDGGILMDVLEFVDLNDYDQIILATDEANGSGTPTTYQIQOQHNVAAQ 81
Qy 87 -----GCDGTS-ASSYGT-----OSPFIQSELETNNVSAFOP-----118
Db 82 PLPYARKFPGGSPSTPIASAFGSVTVNHOAHRAASPLNVGPARNSNDPAPQPHSKERRD 141
Qy 119 --YQATSEIEVTPSVLRANMPLHVRSSD-----PAL-IGLSTSVDSNFSSEEPS 166
Db 142 SVVEVSSFPQIQDSGLRVSTPKPSRQSEVDYDCKPNNQPLRLSLRTEASGR--TEEAT 199
Qy 167 RKNPTKSTTAGFLKONTAGSPKTCRRK--DENYKSLPRDTSNMGNOPORONARSS---221
Db 200 PVKQSHVTLSPYEYKLLAEODEKSKRRKHNDN-----PGRARSGDRKSRLT 248
Qy 222 --LSASHPMVGKWLKOEODEDEDEEDNSRVEPVGHADTGLEHINPFSLDWYKLYEVPN 279
Db 249 DALLDARDRIADQLEQNPAAE--TKSQMIRVK-----IDG--PMPTSLVTPPIPEKSE 301
Qy 280 DGGPGLIHV-----VPFSARGRTGLGLVYKLEKKGAKAHEMLFRENDCIYVINDG 330
Db 302 NEKQILIEVNAVPEDESSELPSTSEPTKLSVQIMKTEDGGRILAKDRIR--VVRASIS 357
Qy 331 DLNRRFEQAQHMFRQAMPTPIHVFVPAANKEQYEQLSQSKNNYSSRFSDQYID 390
Db 358 DL-----AAVTSRPVTLII 371
Qy 391 NRSVNSAGLHTVQARPLNHPBPQIDSHSLRPHSAHPSGKPPSAPASADONVSTVSSG 450
Db 372 NRSLESF-----LEQES-----SAKPIQ--SALQQA-----395
Qy 451 YNFKTKGRLNIDLKKGTBGLSITSRDYITGGSAPYIYKNIPLGAAIQDRILKAGDR 510
Db 396 -NTQYIGHTTVLELIKSSNGFPTVGR--TANGERLFIYGYKPGVAL--GHLKSGDR 451
Qy 511 LIEVNGVDLVGKQOEYVSLRSTKMEGYSLVFR--QEDAFHRELNAE-----559
Db 452 LLETNGTPTGQMTQSEIVEKLEKTMGEEKIKFLVSRYSOSALINSTASSENKNEETLKY 511
Qy 560 -----PSOM--QIPKETAEDEDIVLPDGTREFLFEVPLANDSGSAGLGVSV 605
Db 512 VEDEKIPQKLPLPALMTPPVPKDTPA-----LSPGASNFELVIFPINSAGAGVSL 565
Qy 606 KGNRSKENH---ADLGIYKSIINGGAASKDGRLRYNDOLIAVNGESLCKTQODAMETL 662
Db 566 KAVRSKSSNGSKVDCGIFIKVMHGAAPKEGGLRVDRIYVEDIDLEPLDREKQOAL 635
Qy 663 RRSMTSEGNKGMIOILVARIKCNELKSPG-----SPPGELPIETALDO 709
Db 626 AKKL-----KEVGMISNVRLTISRNEC--NPGQISRLDSITTYDASSPS--671
Qy 710 RERRISHLSYSGI-----EGLDSPSRNAALSRIMGESGKYQLSPYVNMPODDT 758
Db 672 -SRMSSHTAPDSSLSPATRGTSAGSSGADSSHSRQSSAS-----708
Qy 759 VILEDDRPLVLPRLHDQ-----SSSSHDVGVVTADAGTMAAALSDADCSLPDVP 814
Db 709 -----SAVPVPARLTEROSIVSDGTSRND-----SEELPDSAD-----742
Qy 815 VLAFOREGFGROGMSERK-----TKQFSDASOLDVFKTRKSKSMDLGIADETKLN 864
Db 743 --PNNREGLGKRLSKRCHGAADQHIKLPD-----IKHQRONASAP-----784

```

```

QY      865 TVDDQKAGSPSRDVGPSLGLTKSSLESLOTAAVEATYLNDDIPHRRPRLINGRCNES 924
Db      785 TSSYQK- -----SKQPRSSSQNNYKSPMLY- --DLP----- 814
QY      925 FRAAIDKSYDKPAVDDDD-----EGMETLEEDTEESSRSGRESVTA- -- 967
Db      815 TTAAASASTSNQNLDDSDMLNRSQSMESIRPVESTLR-GTQGIPTGSSSKVQFMQAS 873
QY      968 -DQ-----PSHLEKQRMQNGNOEKGDTRKKDKDTGKEKKDKDKDKAKKAKKMLKGLGD 1022
Db      874 PDQHPFPFGALLMLKN-----EESRSDKSR-----KSGMAMRN 910
QY      1023 MFRFGKHKKDKDTIEKTKIKIQESFTSEERIRMKQEDERIQATREFREQANER----- 1078
Db      911 FFGGSGSRDASPKPTQTESVOLRSVERPKSIDERNNGSSERAPPLPPIQSGRGSGG 970
QY      1079 -----DYALIDQPHRFGCDDDELWYGVSYSYSGM-----ALMARPOSREGHMMDA--- 1125
Db      971 NVEVDYGE-----PYGLIPQYPHHTTGTGYESYADSELYDRAAHRHYPHGGPIIDEDEX 1024
QY      1126 LVAQYKPKRPNKSPSPVDS-----NRSTPSNHDRITQIKRQFQQAQKQDEVEDRRRTYSPQ 1181
Db      1025 IYRQOSTSGN---SPINTSSYVNYGLPASN-----AHVGS 1057
QY      1182 PWPNAPRATQSSRHVSAYEVQMGQROERERESSQQAORQSLPPOSRKNASSVSQDSWE 1241
Db      1058 RIP---PQTSSG--SISKTSGAMRRRYPAETIEDYAIHQ--QIPQOSTR----- 1099
QY      1242 QWYSPGEGFQSAKENPRTSSYQGSRNGLYIGCHGFNARVMLET 1283
Db      1100 --YQGGSG--SGRGNDYHHMFNSMFAVYGGAVGAPAIKS 1137

RESULT      4
T46612
multi PDZ domain protein 1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 21-Jul-2000
C:Accession: T46612
R:Ullmer, C.; Schmuck, K.; Figge, A.; Lubbert, H.
FEBS Lett. 444, 63-68, 1998
A:Title: Cloning and characterization of MOPPL1, a novel PDZ domain protein.
A:Reference number: Z23104; MVID:98196865
A:Accession: T46612
A:status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2054 <URL>
A:Cross-references: EMBL:AJ001320; NID:92959978; PIDN:CAA04681.1; PID:92959979
A:Experimental source: brain
C:genetics:
A:Gene: MOPPL1

```

Query Match	5.1%	Score 361.5	DB 2	Length 2054
Best Local Similarity	20.5%	Pred. No. 5e-10		
Matches 246	Conservative 159	Mismatches 433	Indels 363	Gaps 48
Oy	189	KTCGRKK---	DENVGSLPRDTSNNNOFQNDNARSSLSASHPMVGMLEKQEDDEGTE	244
Db	3	ETIDKNNAQAARLQSLKERGDVANEKDSLTKSVLOS--	PLFSQTLSTQSTLQQLKD	60
Oy	245	EDNSRVEPVGHATGTEHPNFS---		267
Db	61	QVNAVATLATANADHA--HTPQSSAIIISMQSESLILSPSNCNLEAISGCPAPAMDGKR		118
Oy	268	---LDDMVKL---VEV---PNDGGPIGIIHVPPRSARGGTTGLLYKRLKRGSKA		312
Db	119	ACEEELDQIKSMAQGRHVEIFELLKPPCCG-LGFSVYGRSSENKGELAIPOEIOESVA		177
Oy	313	EHEHNIFRNCQIYVINGDILRNRRFPQAGHMFRQAMRPELIWFHVVPANKEQYEOLSOS		372
Db	178	HRDGRLKTTDQILAIN-----		193

QY	373	EKNNTYSSRFSDQYIDNRKSVNSAGLTHTVQARRLNHPPEQDISHRLSPSAPHS-GKP	431
Db	194	-----GOVLDQOTITTHOQAISILOKA-KOTIOLVLRKSLPHISSPRISRS	237
QY	432	PSAPASAPQWNESTVSSGVWTKIKGRRLNTLQKKGTEGLFSTSRDVTIGSSAP-IYV	490
Db	238	PSA-----ASVSAHSNPTHQHETIELVNDGSGAFGCI-----IGKATGATYV	282
QY	491	KNLEPRGAAIODRLKAGDRILEVNGCVLVKSOEVEYSLIRSRMEGVSLVFR-OE	548
Db	283	KTILPGVADVADGRLCSGDGHILIKIDTDLAAGSSQVAVQVLR-QCGNRVLTMLARGAVE	340
QY	549	DAFHPRLEI-----NAEPSOMOIPKREKNADEDIYLPDGTREFLEFVPLANDSSAGLG	602
Db	341	ETPPAPSSLGITILSSSTSTSEMRVDASTQKNES-----ETFPVELT-KNVQSLG	388
QY	603	VSVYKGRKSENHADIGIFPKSTIINGAASKODRLRYNDOLIAVNESLIGKNGDAMETL	662
Db	389	ITTAGIYGDKLEPSGIFPKSTIKSSAVELDGRIGQDOIVAVDGTNLGFTNQQAVEL	448
QY	663	RRMSYEGNKRGMQIOLIVARR-ISKCNELKS-PGSPGPELPIETALDDRERISHSLXS	720
Db	449	RHNGQI-----VRLTMRKGAQSAEATFSEBDTAKVDLTAENYEXK-----	490
QY	721	GIGLDESPSRNALSIRINGESKY-----QSPYTMPODDTVIIEEDRL-----PYLP	770
Db	491	-----BESLSLKRSTSLIPREEGYPILSTLEETEDYQEAALTKWQIRIMCINYEIYV	545
QY	771	PHLSDOSSSSSHDVGFTVADACTMAKKAISDSADCSLSPVDVPLVAPQREGGRQSMSE	830
Db	546	AHYSKFESENG-----LGISLEATVGHMFIRSVLPE-GPV-GHSQKLESGDELLIE	593
QY	831	KRTKQFSDASQDLDPVTKRKSMDL-----GLAETKLTNYVDQKAGSPSRDY	878
Db	594	VNGINLIGENHODVNIKLKELPIDYTWCCRRVYPTPALSEVSDLIHDELTREKPHIDL	653
QY	879	GPSLGLKSSSLESLOTAVAETLNGDIPHRPRPRIIRG-----RCGNES-----F	925
Db	654	GEFIG-----SSETEDDPLAMSDVQNAE-EIOTPLAMWEAGIQAIELEKSGRGLGFTILDY	709
QY	926	RAALDKSYDKRA-----VDDDECGMETLEEDYE-E	954
Db	710	QDPID-----PANNVYIYRSLVPGSIAGLXKGRLEPGGRIMEVNDINLENSTLEEAVALK	764
QY	955	SSRSGRESVSTASDOPSHSLEROMNGNOEKDGTKDRKKDKTKGRKKKDRKOKKMMAKK	1014
Db	765	GAPSGMVRIGYAKPLPLSPBEGYVSAKED-----RFLCSPHCKK-----	804
QY	1015	GMKGLGIMFEGFKHRKD-----DKLETKIKIIOESFTSEEBERIRMKOEORIOAKTR	1068
Db	805	---MGLSKALF---RALDALIDIPDAASVARESRESQFSPDINSVYSTQ-----	848
QY	1069	EFFERQARERDYAEIOTDFHRTFGCDDLMYGGVSSYBESMALNMRPOSPEG-----	1120
Db	849	-----ASVLSLH-DGACSDGMNYG-----PSTLSSPPKDYVNTSSDLY	884
QY	1121	---HM-MDALYAO---VKPRNSKP---SPVDSNKRSTPSNHDRIQRLROEQOAK-----Q	1166
Db	885	LGHLHSLEELYQNLLQRGHAGSPPTDMSPATSGFTVSDTPANAVKQKEACATYAMT	944
QY	1167	DEBVEDRRRTYTFEQPWPAPR---ATQSGRISVSEVEMORQOREEERESSQAOARQYS	1223
Db	945	PSOLPSGLSTTELAPALPAPVAPKYLTEQSSSIVDSAEVTLQMSQSEAFERYTTLAKGSS	1004
QY	1224	L 1224	
Db	1005	L 1005	

RESULT	5
T30259	
multiple PDZ domain protein - mouse	
C:Species: Mus musculus (house mouse)	



Thu Jul 25 08:38:40 2002

Db 765 FVNHHDSLNRHHPVLQVRFKLCIYVQNNIESFIIKAIIRQE-----VROQHEVN 814  
 Qy 567 KETKAEDEIVLPDDGRREFLTFEYPLN-----DGSAGLGSV 604  
 Db 815 GYSRSEFSMTSWSPCSTRSVSPGSPISLGSWADVVELPHLERVYKLOKALPLGAV 874  
 Qy 605 VKNRSKKNHADLGIYKSIINGGAASKDQRLRVNDQILAVNGESLIGKTNDAMETLR 664  
 Db 875 LGDKDKGVN--GCYKSLICGKAYALDRIQVGFITKINTESLRNTNSQAAAILKR 931  
 Qy 665 SMSTGKNKGMILIVARRISKNEKLSKPSGPPELPIETALD-----DRERLSHLY 719  
 Db 932 T-----NLYGTF-----CN-----VTYITSDADATKWRKRFQRPSES-- 962  
 Qy 720 SGIEGLDESPSRNALSRIMGESGYQLSP-----TVNPFODT-----VIEEDRLP 767  
 Db 963 -----SSPIINRLSPKVF--PKFYRSPFMQROESQSKTEMTDDETEAPSLMTDS-- 1009  
 Qy 768 VLPPH-----LSDQSSSSSHD-----VGEVTDAGTAKAKAI-- 800  
 Db 1010 -MSHITFDLAEKSSKSHHDEQVNRKSLIDGVEVDDEFNLIKAITDAILRL 1068  
 Qy 801 ---SDSADCS-----LSPVDVPIALPQREDFGOSMEKTRKQFS 837  
 Db 1069 HKTKDMSCKNKRREERLESPLPPPEVLSPPKSPVASVPT--PROLEVYLR--S 1123  
 Qy 838 DASOLDPEVTKRKSMDLGTADETKLTVDQKAGSPSDVQPSL--GLKSSLSLEQL 894  
 Db 1124 TASSLEYHSGQTSQLTLSTEEVLOATPPSPENKSSVSPSISPSIK----- 1174  
 Qy 895 TAAVEYTLNGDIPFRPRRIIRGCGNESRAIDKSYKPAVDDDDGEMETLEEDTEE 954  
 Db 1175 -LAGEVTAEEI-----EYKKEQAEV-----VDPA-ETATGAEVETSTPLAEALG 1219  
 Qy 955 SRSRGSSTASTADPSHLEKQNGNOEKDKTDR--KKDKTKRKKKDRKEDKDKK 1011  
 Db 1220 NSKDD-ESTTSLISQSVLQTOALNSTEVNNSKRVTSKTPSGESLONQARQLVRSK 1278  
 Qy 1012 -----AKGMLKGLGDMF-----RFG 1027  
 Db 1279 YWGERTVTLVREPNKSFGISIVGKRVYQKGLPGTGNVCGIFIKSVLPNSPAGRSK 1338  
 Qy 1028 KHRDDKLEKYGKIKIQESFTSEERIIRMKOEORIOAKTRREFRQARREDAVIEDFH 1087  
 Db 1339 QANNMGDRIVSYNDVLDLDA--THEQAVNAIKNASPYR-----FVLSLH 1381  
 Qy 1088 RTFGCDDELMTGYSYEGSMAL-NARPO-----SPREGHMADALYAOVKRPNRSK 1137  
 Db 1382 TN---QANNINSASNSTVGSVRFENAKPEDELPTALVPLK-PMISSGSGSTKPAHF 1437  
 Qy 1138 PSPVDSNRSTPSNHDRIQLRQEQOAKODEVED--RRRTYFEQFPWPNARPA-- 1189  
 Db 1438 PPSISITTTTS--MESESKKEEATSSPDIORENTVVRKRDVREQPEANPEETS 1493  
 Qy 1190 -----TOSGRHSVSEVQORQROBERSSQAO-----ROYSLPR 1226  
 Db 1494 IYEKTPPKRIKAKSSSKESNDKREIKOKSYEREKSVSKSVSIKHHQESTIIR 1553  
 Qy 1227 QSRKN-----ASSVSQDSMGNQNSPGEGFOS 1252  
 Db 1554 KSPSNETAPLIVSDVSSSTHEDEPQAMSPSTSFDT 1589

A:Reference number: I53483; MUID:94116679  
 A:Accession: I67629  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2466 <RES>  
 A:Cross-references: GB: D21210; NID: 9452191; PIDN: BAA04751.1; PID: 9452192  
 C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GIGF domain homolog <B41>  
 F:574-868/domain: protein 4.1 membrane-binding domain homolog <GIG2>  
 F:1354-1430/domain: GIGF domain homology  
 F:2218-2437/domain: protein-tyrosine-phosphatase homology <PMP>  
 Query Match 4.7%; Score 332; DB 2; Length 2466;  
 Best local similarity 20.9%; Pred. No. 2, 1e-08; Indels 244; Gaps 32;  
 Matches 193; Conservative 134; Mismatches 352;  
 Qy 80 DEQDPHGGDGTSSASTGTQSPETIGSELG-----TNNVSAFOPQATSEI--E 126  
 Db 879 DAQDI-----ERASFRLNLQAESVKGFMKGRALISGSLASTLKLAVRPLSVQABILKR 934  
 Qy 127 VTPSVLAAMPPLH--VRRSSDPAL-----IGLTSVSDSNFSSPEPRKPTFMSTTAGF 179  
 Db 935 LSCSELSTYPLONSSKKNKDKASWEKPREMSKSYHLSQASLYPHRKN----- 984  
 Qy 180 LKONTAGSPKTCRKKDENTRSLPR-DTSMGNQFORNARSSLSASHPVGYKLEKQEQ 238  
 Db 965 VIVNMEPPQTVAAELVGKPSHOKSRDAESLAGVTKLNKSKSVASLN-----RSP 1034  
 Qy 239 DEDGTEEDNSRVEPVHADTGLEHIFNEFLDDM-----VKLVPPNDGG-PL 284  
 Db 1035 EKKRHESSSSLEDDGQA-----YVLDVLLKMSIYSSPERITLVNLKKAQYGL 1085  
 Qy 285 GIHVVPFSARGRTGLLVKRLKRGKAEHLENRENDCTIVRINDGDLANRRFEQOAHMF 344  
 Db 1086 GFQIIGERMGRLDIGITISSVAPGGPADLGLCPGDRILIVSNVSLSEGVSHHAILEL 1145  
 Qy 345 ROA-----MKPTIWFHY-----VPAANKQY----- 366  
 Db 1146 QNPEDVTIVISQPKRKISKVSTPYHLTNEMKNYMKSSYMODSADSSSKDHMSRGT 1205  
 Qy 367 -----EQLSOKERNYSSRF-----SPDS 386  
 Db 1206 LRHISENSPSPSGGLRGSLSQSQRTEASLSQSVNGFASHLQGOQWQESQHSPPSP 1265  
 Qy 387 QYIDNRSVNSAGLHTVQ--RAPLNHPPEQID-SHSRLPHSAPSGKPPASAPARQNY 442  
 Db 1266 SVISKATEKETFTDSNOSTKTKRGIDVTDYSDRGSDMDDEATYSSODHOTPKQESSSS 1325  
 Qy 443 PSTVSSGYNT-----KRGKRLNLQAKCTBELGFSIT--SRDVTIGSGAPVYKNIL 494  
 Db 1326 VNTSKMNFKTFSSSPKPGDIFEEVLAKNDNLGISTVGVNTSVRHGG--IYVKAVI 1382  
 Qy 495 PRGAIQGRKAKGDRILEVNGVDLVGKSOEYVSLRSTKMGCTVSLVFRQEDAFHPR 554  
 Db 1383 PQGAASDSDRIHKGDRIYALVNGVSLGEGATHQKQAEVLRNT--GOVYHLLKKGQSPYSK 1439  
 Qy 555 E-----LNAEFSQKQIP-----KETKAEDEVILTLPQTRFELTFEYPL-NDSGSAG 600  
 Db 1440 EHVPTPOCLSDPQANQAGGPEKAKTTOVDYGV-----TEENFEFKLEKNSSGIG 1493  
 Qy 601 LGVSVKGNRSKKNHADLGIYKSIINGGAASKDQRLRVNDQILAVNGESLIGKTNDAME 660  
 Db 1494 FFSREBNLIPEDILMSIYVVKLIFPGQPAESGKIDVGIYIAKYNAGSLKLSQCEVVIS 1553  
 Qy 661 TLRSMSTEGKNKGMILIVARRISKNEKLSKPSGPPELPIETALDNRERRISHLYS 720  
 Db 1554 ALKGTAPF-----VFLILCR-----PPPGVLPPIETAL-----LTPLOS 1587  
 Qy 721 GIEGLDESPSRNALSR-----WGESGYQL--SPYVNMPODVTIIEEDRLPVL 769  
 Db 1588 PQAVLPNSKDSQSPCEQDSTSDENEMSKDKKQCKSPSRDSDSSGSGEDDLVTA 1647  
 Qy 770 PPHLSQDS-SSSHDDVGFVADAGTWAKAALS-----DSADC 806

RESULT 7  
 I67629  
 protein tyrosine phosphatase (PTP-BAS, type 2) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 16-Jul-1999  
 C:Accession: I67629  
 R:Maekawa, K.; Imagawa, N.; Nagamatsu, M.; Harada, S.  
 FEBS Lett. 337, 200-206, 1994  
 A:Title: Molecular cloning of a novel protein-tyrosine phosphatase containing a membrane



T15617

hypothetical protein C25F6.2 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

R:Bentley, D.  
 Submitted to the EMBL data library, October 1995

A:Description: The sequence of *C. elegans* cosmid C25F6.

A:Reference number: Z18377

A:Accession: T15617

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1131 <BEN>

A:Cross-references: EMBL:U39742; NID:g1049455; PID:g1049459; PIDN:AAA0434.1; CESP:C25F6

A:Gene: CESP:C25F6.2

A:Introns: 105/1; 210/3; 283/3; 316/1; 346/2; 463/3; 566/1; 722/2; 897/1; 991/2; 1032/1;

## Query Match

Best Local Similarity 22.5%; Pred. No. 2e-08;  
 Matches 176; Conservative 105; Mismatches 304; Indels 198; Gaps 29;

Db 817 GSRQSLSFKRKP-----FVKSTDRL-----NDLNESSVAEEPV-----WSYQAVE 859  
 QY 802 DSA 804  
 Db 860 QQA 862

## RESULT 10

A54971

protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 13 - human

N:Alternate names: Fas-associated phosphatase FAP-1; protein-tyrosine-phosphatase hPr

C:Species: *Homo sapiens* (man)  
 C:Date: 11-Nov-1994 #sequence\_revision 08-Feb-1996 #text\_change 01-Dec-2000

C:Accession: A54971; A55114; I55955; I53483; S46955

R:Banville, D.; Ahmad, S.; Stocco, R.; Shen, S.H.

J. Biol. Chem. 269, 22320-22327, 1994

A:Title: A novel protein-tyrosine phosphatase with homology to both the cytoskeletal

A:Reference number: A54971; MUID:94350588

A:Accession: A54971

A:Molecule type: mRNA

A:Residues: 1-2490 <BAN>

A:Cross-references: GB:U12128

A:Note: sequence shown follows authors' translation at positions 62-63

J. Biol. Chem. 269, 24082-24089, 1994

A:Title: Cloning and characterization of PTP1, a protein tyrosine phosphatase with s

A:Reference number: A55114; MUID:95014139

A:Accession: A55114

A:Molecule type: mRNA

A:Residues: 1-61, 'GS', 64-839, 'D', 841-1055, 1075-1133, 'FH', 1136-1210, 'T', 1212-1383, 1389

A:Cross-references: GB:X60289; NID:9515030; PIDN:CA56563.1; PID:9515031

A:Title: T: Title, S: Kitada, S.; Reed, J.C.

Science 268, 411-415, 1995

A:Title: FAP-1: a protein tyrosine phosphatase that associates with Fas.

A:Reference number: I59595; MUID:95323528

A:Accession: I59595

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1279-1888 <RES>

A:Cross-references: GB:I34583; NID:9806291; PIDN:AA041755.1; PID:9806292

R:MacKawa, K.; Imagawa, N.; Nagamatsu, M.; Harada, S.

FEBS Lett. 337, 200-206, 1994

A:Title: Molecular cloning of a novel protein-tyrosine phosphatase containing a membr

A:Reference number: I53483; MUID:94116679

A:Accession: I53483

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-61, 'GS', 64-839, 'D', 841-1210, 'T', 1212-1383, 1389-2299, 'QW', 2302-2490 <RE2

A:Cross-references: GB:U12109; NID:9452189; PIDN:BAA04750.1; PID:9452190

C:Gene: GDB:PTPN13

A:Cross-references: GDB:306348; OMIM:600267

A:Map position: 4q21.3-4q21.3

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GIGF domain homolog

F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1099-1175/Domain: GIGF domain homology <GIG1>

F:1373-1454/Domain: GIGF domain homology <GIG2>

F:1799-1870/Domain: GIGF domain homology <GIG3>

F:1893-1967/Domain: GIGF domain homology <GIG4>

F:2242-2461/Domain: GIGF domain homology <GIG5>

F:2413/Active site: Cys (phosphotyrosine intermediate) #status predicted

F:2419/Blinding site: substrate phosphate (Arg) #status predicted

## Query Match

Best Local Similarity 4.6%; Score 321; DB 1; Length 2490;  
 Matches 196; Conservative 128; Mismatches 364; Indels 250; Gaps 32;

QY 80 DEORPHGSGTASSTGQSPETFGSELG-----TNNVSARQPOQATSEI-E 126

Db 879 DAQDI-----ERASFSLNDQAESVRCFNGRAITGSLASTINKLAVRPLSVQAEIIRK 934

```

QY 127 VTPSVLRANMPLH--VRSSDPAL-----IGLSTVSDSNFSSEPSRRKNPTFMSTTAGF 179
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 935 LSCSELSTLQPLONSSKEKNKAMWEKPREMSKSYHDSLSQSLYHPRKN-----984
QY 180 LKONTAGSKTCDRKDEKDEYRSLPR-DTSMNSQFORDNARSSLSLHPWGMLEKQED 238
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 985 VLVNNEPPPTVAELVKGPSHQMSRSDASLAGVTKLNSKSVASLN-----RSP 1034
QY 239 DEDGTFEDNSRVEPVGA-----DTGLEHLPNFSLDDM-----VKLV 275
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1035 ERKKEHSDSSSTIEDQOAVLGMTMSSGNSSQVPLKENDVLHKMWSIVSSPERETITLV 1094
QY 276 EYVPNDGG-PLGIHVYFPFSARGRTIGLLVRLKLEKGAHEMLFRENDCIVRLNDGLRN 334
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1095 NLKRDAKYGLGFQIIGGEMKGRLDIGIFISSVAPGCPADLDGLKPGDRILISVNSVLSG 1154
QY 335 RFEQAOAHMFRQA-----MRTPIWHVYPA-----NKQYEQ-----368
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1155 VSHHAILEILQNAPEEDVTLVISQPKERISKVPSTPVHLNEMKNYMKSSYMODSAMDSS 1214
QY 369 -----LSQSEKNNYSSRF-----382
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1215 SKDHHSKGTLLHISENSFGPSGGLREGSLSDSODRTESASLSQSOVNGFFASHLDQTM 1274
QY 383 -----SPDSQYIDNRSVNSAGLHTVQ---RAPRLNHPEPID-SHSRLPHSAHPGKBP 432
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1275 QRSQHSFSPSVISKATEKETFTDSNOSKTKKPGISDVTDYSDRGSDMEATYSSSDQH 1334
QY 433 SAPASAPQWVFSTVYSSGNT-----KTKGRNLNIOLEKGTGEGSIT-----SRD 479
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1335 QTPKQESSSVNTSMKNKFTSSSPKPGDIFEVELAKNDNSLGISTVLEPKDGGVNS 1394
QY 480 VTIGSAPITVYKNILPRGAIDGRKAGDRILEVNGVDLVKSGOEYVSLRSTMEGT 539
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1395 VNHGG---IYKAVITPQGAESDGRTHKGDRIYAVNGVSELEGTTHQAVETLANT---GQ 1448
QY 540 VSLVFRQDAFHPR-----LNAEPSQMLP-----KETKADEDIVLTPDGTRE 586
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1449 VVHLLEKQSPSKHEHVPTQCTLSQNAQOGGPEKVKTTQVVDYSFV-----TER 1502
QY 587 LFEFVPL-NDSSAGLGVSVKGRSKENHADLGIFFKSLIINGAAKDRRLVNDOLIV 645
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1503 NTFEYKLFENSSGLGFSFREDNLPEQINASTIVRKKLFPQPAAESGIDVGLVLY 1562
QY 646 NCESLGKTNODAMETLRSKSMTEGKRMQIOLIVARRISKCNELKSPSPGPELPIET 705
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1563 NGASLKLGSQEVISALRGTAPE-----VFLLCR-----PPGVLEPIDT 1603
QY 706 ALDNERRISHSLYSGIEGLDPSRNALSLRI-----WGESGKYQL-SPTVNNP 754
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1604 AL-----LTPLOSPAQVLPNNSKDSQSCVEOSTSSDENEMSKSKQCKSPSRDS 1656
QY 755 QDDVTIIEDRLPVLRPHISDOS-SSSHDDVGVYADAGTAKAALIS-----801
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1657 YSDSSGSGEDDLVTAANISNSTWSALHOTLSNMYSOASHHEAPKSOEDTICTMFTY 1716
QY 802 -----DSADCSLSPVDVPLVAFQREGFGROGMS 829
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1717 QKIPNKEFEEDSNPSPLPPDMAFGSYQPSSESASSS 1754

```

```

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2294 <RES>
A:Cross-references: GB:D21211; NID:9452193; PIDDN:BA04752.1; PUD:9452194
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain
F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>
F:1182-1258/Domain: GLGF domain homology <GLG2>
F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

Query Match 4.5%; Score 314; DB 2; Length 2294;
Best local Similarity 20.3%; Pred.No.1.5e-07;
Matches 201; Conservative 143; Mismatches 360; Indels 288; Gaps 38;

QY 25 SLIQAVTRYRKAIADPNVMIQVHRLHEDGGLDLDLTL-----CDVADKDRIVAAVF 79
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 668 SLIQHTLTCHQ-----YYLDLRK-----DLBERMNC-DKSLSLASL 705
QY 80 DEODPHHGDDTGSASTGTGSPETFGSELGTNNVSAFQ-----PYQATSEIETPSYLRA 134
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 706 ALQ-----AEYGDYQPEVHG-----VSFMRMHHYLPARMKELDL-SYIKE 745
QY 135 NMF-LHYRRSSDPALLGLSTVSQD-----SNFSEPSRRKNPT---RKSTT 176
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 746 ELPRKLH-----NTYVGASEKETELFELKYCQRLTEYGVHFNHVEPKSQGTILLGVCS 799
QY 177 AGFLKONTAGSPKTC-----DRKK---DENYRSLPRDTSNMSNOFORDNARSSLSASH 226
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 800 KGVLYEYVNHGVRTVLRFPMWRETKKISPKKKITTLQNTSDGILKHGQVNTSKICQYLH 859
QY 227 -----PVMKMLEKQODEDGTEDNSRVEPVGADTGLEHLPNFSLDMKLYE 276
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 860 LGSYQHKFQLQMRAROSNOADIDVLLHKRMSIYS-----SPEREITLVN 904
QY 277 VPPNDGG-PLGIHVYFPFSARGRTIGLLVRLKLEKGAHEMLFRENDCIVRLNDGLRN 335
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 905 LKRDATYGLGFQIIGGEMKGRLDIGIFISSVAPGCPADLDGLKPGDRILISVNSVLSG 964
QY 336 RFEQAOAHMFRQA-----MRTPIWHVY-----VPAAN 362
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 965 SHHAILEILQNAPEEDVTLVISQPKERISKVPSTPVHLNEMKNYMKSSYMODSAMDSS 1024
QY 363 KEQY-----EOLSEKNNYSSRF-----382
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1025 KDHHSKGTLLHISENSFGPSGGLREGSLSDSODRTESASLSQSOVNGFFASHLDQTM 1084
QY 383 -----SPDSQYIDNRSVNSAGLHTVQ---RAPRLNHPEPID-SHSRLPHSAHPGKBP 433
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1085 ESQHSFSPSVISKATEKETFTDSNOSKTKKPGISDVTDYSDRGSDMEATYSSSDQH 1144
QY 434 APASAPQWVFSTVYSSGNT-----KTKGRNLNIOLEKGTGEGSIT---SRDVTIGS 485
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1145 TPKQESSSVNTSMKNKFTSSSPKPGDIFEVELAKNDNSLGISTVLEPKDGGVNS 1203
QY 486 APIYVKNILPRGAIDGRKAGDRILEVNGVDLVKSGOEYVSLRSTMEGTSLVF 545
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1204 --IYKAVITPQGAESDGRTHKGDRIYAVNGVSELEGTTHQAVETLANT---QCVALL 1258
QY 546 RQDADFHPR-----LNAEPSQMLP-----KETKADEDIVLTPDGTREIFTEPV 592
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1259 EKQSPSKHEHVPTQCTLSQNAQOGGPEKVKTTQVVDYSFV-----TEENFEV 1312
QY 593 L-NDSSAGLGVSVKGRSKENHADLGIFFKSLIINGAAKDRRLVNDOLIVNGESL 651
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1313 LFNKSSGLGFSFREDNLPEQINASTIVRKKLFPQPAAESGIDVGLVLYAVNGASL 1372
QY 652 GKTNODAMETLRSKSMTEGKRMQIOLIVARRISKCNELKSPSPGPELPIETALDRE 711
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1373 GLSQQEVISALRGTAPE-----VFLLCR-----PPGVLEPIDTAL-----1409
QY 712 RRISSHLYSGIEGLDPSRNALSLRI-----WGESGKYQL-SPTVNNPQDDTVI 760
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1410 --LTPLOSPAQVLPNNSKDSQSCVEOSTSSDENEMSKSKQCKSPSRDSYSDSSG 1466

```

QY 761 IEDRLPVLPHLSDQS-SSSHDDVGFVTADAGTWAKAAS----- 801  
 Db 1467 SGEIDLVTAPANISNTSMSSALHQTLSNMVSOQASHHEAFKSOEDTICTMYFPQIKPK 1526  
 QY 802 ----DSADCSLPDYDVLAFOREGREGOSMS 829  
 Db 1527 PEFEDSNPSPPLPDMAGQSYOPQSSASSS 1558

## RESULT 12

T13564  
 microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)  
 N:Alternate names: hypothetical protein EG:49E4.1  
 C:Species: Drosophila melanogaster  
 C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
 C:Accession: T13564  
 R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.  
 Submitted to the EMBL Data Library, April 1999  
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
 A:Reference number: 217689  
 A:Accession: T13564  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-5327 <SPA>  
 A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1  
 C:Genetics:  
 A:Cross-references: Flybase:FBgn0025392  
 A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1  
 A>Note: EG:49E4.1  
 C:Superfamily: Drosophila 576k microtubule-associated protein homolog

Query Match 4.2%; Score 296.5; DB 2; Length 5327;  
 Best Local Similarity 18.6%; Pred. No. 3 8e-06;  
 Matches 279; Conservative 227; Mismatches 569; Indels 423; Gaps 60;

QY 91 TSASSTGQSPETFGSELGNNVSAFOYQANSEIEVTPSVLR-----ANMPLHVRSS 144  
 Db 693 SSPESTAKAK-----EANNKRVLESKQOARVQATSTVRRTVSTASERVOQQA 746  
 QY 145 DPALIGLSTVSDSNFSSEPSRKNPTRWSTAGF-LKONTAGSKTCDRKK----- 195  
 Db 747 KTAATG-AVQATQRPKIPSRPGVSPSKRAPPGSPVAKO-AKKRADLKKTRLDKGT 802  
 QY 196 -DENYSLP-ROTSNMSNFOFQDNASSISASHPVYQWLEKQEDDEGTEDN-SRVE 251  
 Db 803 TDSLSLVSTPSADATAAKLDLTASOELDA-----EKORLDDLKEQEVAREIE 853  
 QY 252 PVGHADTGLEHLPNFSLDDM-----VKLVEYPNCGPLGIHVVPFSARGRTGL 302  
 Db 854 AV-----FSRDEMKRQHQOQKAKELREMPAEGTDG----- 884  
 QY 303 VKRLKGGKAHEHNFENDCIVKINDGDLNNRRPEQAOHMFROAMTPILMFHVYPAAN 362  
 Db 885 -----ENEDDEBEYLILKEVEEQYTEDSI-----VEQESSM----- 918  
 QY 363 KEQEQQLSQSEKNNYSSRFSPDSQYID-----NRVSNAGLHTVORARLNHPPE 413  
 Db 919 KEEELQKHQDSQSEKKRKSAEELIEALAYEAERKARLEGASARDESELDVEPE 978  
 QY 414 QID-----SHSRPLHS-AHPSGKPPSPA----- 436  
 Db 979 QSKIKAEVQDIATAKDIASRTEQOLAKPAEELSSPPEEKLTKSTDTDDQIGAVY 1038  
 QY 437 -----SAPONVESTVYSSGYNT-----KKIKRNLIO-----LKKGTEGIG 472  
 Db 1039 DVLFPVNLQESLPREKFSATIESGATTPILPDEERIPLDQIKEDLVTEKRYAKETKBAE 1098  
 QY 473 FSTSRDYVTGGASPIYVKNILPRGA-AIDGRILKA-----GDRIL-----EVN 515  
 Db 1099 AIVVATVQTLPEAAPIADITLILASATKADAPKADANAEALGELPDSERVLPMKMTPEAQON 1158

QY 516 GVDLYGKSOEYV-----VSLRSTKMEGTSLVFRQEDAFHPR----- 554  
 Db 1159 LLRVITKTPDEVDLPRVHEADLGLYKDSQDANKSISHKESKKEKEFTDDEKENVG 1218  
 QY 555 -ELMAEPSQMP-----KETKAEDEDLYLPDGT-----REFLFEVPLNDGSAIGY 603  
 Db 1219 EIELGDEPKNVDSHLVLKESVOEVAEKVYLETIVKROEIVEATVITQENQEDLME 1278  
 QY 604 SVKGNRSKENHADLGIYKVSIIINGAASKDGR-----LRVNDOLIA-----VAGESILG 652  
 Db 1279 QVAKDEEHQKTESGITTEKKAQKASPTPEKETSIDSDELTAQOLADPTVPPKSAKD 1338  
 QY 653 KTNQDAME-----TLRRSMTEGKRMQIDLIYARRISCNELKSPGCPGELPTETALD 709  
 Db 1339 REDTGSIESPPIIEALIEVE-----VOAKQEKQKP--VPAPAEAIKT----- 1378  
 QY 710 RRRRLSHSLYSGIELDPSRNALSRINGESG-----KYOLSPVYNNPQDDTVY 760  
 Db 1379 -----EKSPILASKETSPESATGVSVEDETEQTKSKSPVSPPESEA-- 1420  
 QY 761 IEDRLPVLPHLSDQS-----SSSHDDVGFVTADAGTWAKAASIDSADCSL 808  
 Db 1421 -KDKSPFASGEAPRPESVASEVDEAGKASRRESIAKTHKDESSLDAKAKQESRRESL 1479  
 QY 809 SPDDVPVLAFOREGFGQSMSEKTKQFSDASOLDPVKTRKSK-SMDIGIADETKLVTV 866  
 Db 1480 AESIKP-----ESGIDEKSAKASKASRPESV-IDKSKESRRESIESLKAEST 1528  
 QY 867 DDQKAGSPSRDVG-----PSLGKSSSLESLQTA 896  
 Db 1529 KDEKSAAPPSEKASRPQSVASVDETEKESKESRRESIASAKPPIEFREVSREPISD 1588  
 QY 897 VAEVTLNGDLPTRPRRIIRGRCNESFR-AAIDKSYDKPAVDDEGNETLEEDIE 954  
 Db 1589 IKDESA-----KPSRSDPSPLASKASRPESVLESKVDEPIKTEKSRRESVAESFKA 1641  
 QY 955 SRSRGRESVSTASD--QPSHSLERQNMQEGKDDTKRKKDKYKERRKDKKDKMKA 1012  
 Db 1642 DSTKDEKSLTSDISRPASVAYENWDAFK--ETSRPESAVGSMKDESKSPSRRES 1698  
 QY 1013 KR-----GMLKGLGDMFRFGKHKDKDKIKTKIKTIOEFTSTSEERIR 1055  
 Db 1699 VADGAAQSRETSRPASVASEAKDADLKLRSRPSITQSKESASIDKESPLASEPESR 1758  
 QY 1056 MKQEOERIAQKREPREQAREROYAEIQDHRFTFGDDDELMTGVSYSYGSMALNAPQ 1115  
 Db 1759 PASVAESVKDEAKESKESRRE-SVAE-----KSPILPSKEAS-----RPA 1797  
 QY 1116 SPREGHMADALYAQVKKPRNS--KPSPVDSNR-STPSNHDRIQLRQEOPOAKODEVD 1172  
 Db 1798 SVAESIKDEAKESKESRRESVAESKSPILPSKEASRPAS--VASIDDEAKESKES--ES 1851  
 QY 1173 RRRYTSFROPMPN--ARPATQSGRHSVYEQMQRQROEERSSQAOAROYSSILPROSR 1229  
 Db 1852 RRESVAESKSPILPSKEASRPAS--SVAESIKDEAKESKES--ESRRESVAESKSPILPSKA 1904  
 QY 1220 KNASSVSO--DSWEQNTSPGEGFQSAKENPRYSYQSGSNVYGLGHGNAYMLETOL 1286  
 Db 1905 SRPASVAESIKDEAKESKESRRESVAESKSP-LPSKEASRPASV-----ASIDDEAK 1957  
 QY 1287 LRQORRKEQOMKOPSPSEBPNYDSY-KKVQDPTAPPKGPFRODV--PPSPQOVA 1340  
 Db 1958 SKESRRESVAESKSPILPSKEASRPASVASEIKDEAKESKESRRESVAESKSPILPSKA 2015

## RESULT 13

138756  
 homolog of Drosophila discs large protein, isoform 2 - human  
 C:Species: Homo sapiens (man)  
 C>Date: 01-Mar-1996 #sequence\_revision 01-Mar-1996 #text\_change 21-Jan-2000  
 C:Accession: I38756  
 R:Lee, R.A.; Marfaglia, S.M.; Branton, D.; Chishti, A.H.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994





QY 475 ITS--RDVTTGGSAPIYKNIILPFGAIIODGRILKAGDRIEYNGVDLVKSGQEEVYSLR 532  
 Db 217 IAGVGNGOHIPGNSIYVFKTIIEGGAHMDGRIDIGDKLAVNSVGLDVAHEDVAVALK 276  
 QY 533 STKMEGVSLVFRQ-----EDAFHREIMAPESQOMIPEKTK-----AEDEDVLVLPDG 562  
 Db 277 NT--YDVVYLKAVKPSNAVLSDSYAPPDITTSYSQ--HIDNEISHSYGLGDYPIAMPTPS 333  
 QY 583 TREELTF-----EVP-----LNDGSGAGLVGSVGNKSKNHADLGIPIKSIINGG 628  
 Db 334 PRRTSVANOLLGEEDIDPREPRIYIHRGSGTGLGFNYIGEDGE-----GIFISFLIAGG 388  
 QY 629 AASKDGRLEVNDDLIJAVNGESLIGKTNDAMETIRSMSTEGNKGMIOLITARRISKON 688  
 Db 389 PADLSELKRGDQILISVNGVDLRNASHQAALAKNA-----GGYVITIAQ----- 434  
 QY 689 ELKSGSPGPELPIETALDD--RERRISHSYSGIEGLDESPRNALSRINGESGKYQ- 746  
 Db 435 -----YKPEYSRFEAKIHDLRQILNSSLGSGTASLNSPKRGFYI--RALFDYDKTRD 487  
 QY 747 ---LSPYVNMPODDVTIIEEDRLPVLPHLSDS-----SSSHDDVGVFTR-----D 791  
 Db 488 CGFLSQALSPFEGDYLVID-----ASDEEMWQARRVHSDSETDIDGIFPSKRKRYE 538  
 QY 792 AGTAKKAAISDSADCSLSPD--VDPVLAFO 819  
 Db 539 RREMSRLKANDWGSSSGQREDSVLSTYE 567

RESULT 15  
 T43775  
 neurabin - rat  
 N:Alternate names: actin-binding protein  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text change 11-May-2000  
 R:Nakamishi, H.; Ohsashi, H.; Satoh, A.; Wada, M.; Mandai, K.; Sato, K.; Nishioke, H.;  
 J. Cell Biol. 139, 951-961, 1997  
 A:Title: Neurabin: a novel neutral tissue-specific actin filament-binding protein involve  
 A:Reference number: 22384; MUID:9801926  
 A:Accession: T43275  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1095 <NAK>  
 A:Cross-references: EMBL:U72994; NID:g2623756; PID:g2623757; PIDN:AAC53454.1  
 A:Experimental source: Brain

Query Match 4.2% Score 293; DB 2; Length 1095;  
 Best Local Similarity 19.5%; Pred. No. 6.1e-07;  
 Matches 269; Conservative 163; Mismatches 477; Indels 470; Gaps 56;

QY 75 LVAVPEODPHHGDSSTGTSPTIFGSELGTN---NV---SAPQYQATSEIE 126  
 Db 30 LKSTPDKPRP---DGEOKTKEGSGQSGRKATGNSVNRKILQMGMENPENAIIA 85  
 QY 127 VV-----PSVLKRNAM-PLHVRSSDPALIGLSTVSQ--SNFS--EES--RKN 169  
 Db 86 KTRGGRSSPQKRRKREVEKTDGVSVKLESSVERISRFDTMDGFSYAKTETRK 145  
 QY 170 PTRMTGTGFLKONTAGSPKTCRRKDKENRSLPRD-----TSMNSNQQRNMASSISA 224  
 Db 146 FERSHESG---QNNRHS---KKEKAGEAPQDEMGSGSKNSGSSLSLSPRTA 197  
 QY 225 SHPWYKMLEKQEDDED-----GTEEDNSRVEPVGHADTGLEHIFNFSLIDMWKLVAYPN 279  
 Db 198 VSPVSOISAVFENSEPGALITPKKAEANSYVTHYPLNTPSYVTMID----- 247  
 QY 280 DGGFLGIHVVPFARGRTLGILVKRLKGAHEHNLIFRENDCLIVRIINDDLANKRPFQ 339  
 Db 248 -----TFG-----RKDSNSRPSNNKQ 264

QY 340 A-----QHMERQAMRPIIWEHVVPANKKEQYEDLQSEKNNYSSRFPDSQYIDNRGVN 395  
 Db 265 ATDTEEPKSAVAVVPEAKQGTSLASLPSEERQLSTAEADVTAPDPTDSDKQSGEP 324  
 QY 396 SAGLHTYQVAPRLNHPPEQID-----SHSLRPLSAPSG-----KPPSAPASA----- 438  
 Db 325 SAESQAMPKSNLTLSRPEPLEDAEAVNVGSEAEQDQRDLTGEGDGLTSPDASASSCKREV 384  
 QY 439 --PQNVSTT---VSSGVNKKIKRNLNLQ--KKGTGELGFSITSRDVTIGGSAPIYKN 492  
 Db 385 PEDSNSEFSGHYMHSDYVAVTRSRYSNDWEGCTE----- 421  
 QY 493 ILRGAALIDGRILKAGDRIEYNGV-----DLVKGSGQEEVYSLMSTMEGTVSLVF 545  
 Db 422 -----QD-----EGDSDENNYVQPMESRLVGLPQEBELIPANKIKF--SCAPIKVF 468  
 QY 546 R--QEDAFHREIMAPESQOMIPEKTKAEDDVLVLPDGTREFLTFEVLPLNDGSA----- 599  
 Db 469 NITYSNEDYDRNDVDVPAASAEYELEKREKLELFP-----VELEKDEGLGIST 520  
 QY 600 GLGVGVKNGRKENHADIGIFVKSIIINGGAASKDGRILRVNDOLIAVNGESLIGKTNDAM 659  
 Db 521 GMGVADAGLEK-----LGIFVKTVTEGGAORDGHIOYNDQIVEVDGISTLVGTQNPAA 575  
 QY 660 ETLRSMSTEGNKGMIOLIVARRISKCNELKSPGSPGPELPIETALDD--RERRISHS 717  
 Db 576 TVLR-----NRKNVRYIGR-----EKPGVSEVALQISQTLQGRRORELLER 620  
 QY 718 LYSIEGLDESPRNALSRINGESGKYQLSPTVNMPODDVTIIEEDRLPVLPHLSDS 777  
 Db 621 HYMOYADDD-----ETGEYATDEE-----EDEGLTLP----- 649  
 QY 778 SSSSHDDVGVTAAGTAKKAA--ISDSADCSLSPDVVLAFOREGFGMSSEKRTKO 835  
 Db 650 -----GGDAIEVFELPENDEMFSPSDLD-----TSKLSKH 680  
 QY 836 FSDASQLDVFYKTRKSKMDLGIADETKLTNYDDQKAGSPRDVPSGLKSSLSLDT 895  
 Db 681 FKEL-QIKNAVTE-----AEIOKLKT-----KIDA 704  
 QY 896 AVAEYTLNCDIFPHRPRRIIRGRGNESFRAAIDSKYKPAVDDDECMETLEEDTERS 955  
 Db 705 AE-----NEKYRWELEKQLOQNTENKEMVXL----- 733  
 QY 956 SRSGRESVSTADQPSHSLERQMNQNEKGDKTDRKKDKTKG----- 997  
 Db 734 -----ESTWIEAQTLCHTVNEHLKETQSOYQALEKKYNKAKKLINDFOKELDPIRROEV 788  
 QY 998 EKKRDROKEDKKAKKAGMLKGLDM---FRP-----GKHKKD 1032  
 Db 789 ERKLEEVERKAHLVEVOGLQVIRIDLEAVFLLKONGTQVNNNNNIFERRRPSGEVSKG 848  
 QY 1033 DKIE--KTGKIKIOESFTS-----EEERIMKQEOBRIOAKTREFERRQAREROVAEI 1083  
 Db 849 DTMEENVEVKGQTSQDGLSDLDLEANVPETERLDSKALKTRAOISVKNRQRPTRR----- 903  
 QY 1084 QDPHRTFGDDDELMTGCVSSYSGSMALNAR-----PQSPREGMMALYLAQYKPR 1134  
 Db 904 -----LVDSVSTDGEDSLERKNFTFNDPFSSTSSAD--LSGLAEKPKTG 949  
 QY 1135 NKPFPSEVDSNRSTPSNHRIORLROFOQAKODVEDVDRRRYTSFQOPPNARPAPOSGR 1194  
 Db 950 LQSOALSDESL---DMT-----DDEITLD-----GQSPK 977  
 QY 1195 HSYVAYEQMORQOBERSSQAOAROYSSLPQSRKNASSVSQDSMEQ-----NYSQK 1248  
 Db 978 HTQS-----QSRVHEMSVOQVSHMLVGLSLDQYVSEPSAONIGEDLQDGLKTKAL 1031  
 QY 1249 GFQAKENPRYSYQSGSRNGYTLGCHGNARVHLEFQELLROQRKKEQO--MKROPSE 1305  
 Db 1032 GMTSSODRALVKKIKAKMKSLT---KARKQOEKQERKLRKEQEBQMQRRKSKSE 1086

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: July 24, 2002, 14:29:54 ; Search time 24.98 seconds

(without alignments)  
2101.830 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 7035  
Sequence: 1 MKYVCFGRTRVYVPCGDGH.....SOVARLRNRQTPKGRPFYS 1356

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	324.5	4.6	2485	1 PTND_HUMAN	Q12923 homo sapien
2	296	4.2	1816	1 AFG_HUMAN	P55196 homo sapien
3	294	4.2	849	1 DLG3_MOUSE	P70175 mus musculu
4	293	4.2	767	1 DLG4_HUMAN	P78352 homo sapien
5	293	4.2	849	1 DLG3_RAT	O62936 rattus norv
6	293	4.2	1095	1 NEB1_RAT	O35667 rattus norv
7	291	4.1	724	1 DLG4_MOUSE	O62108 mus musculu
8	286.5	4.1	904	1 DLG1_HUMAN	Q12959 homo sapien
9	286	4.1	724	1 DLG4_RAT	P31016 rattus norv
10	286	4.1	870	1 DLG2_HUMAN	Q15700 homo sapien
11	284.5	4.0	817	1 DLG3_HUMAN	O92796 homo sapien
12	279	4.0	852	1 DLG2_RAT	O63622 rattus norv
13	278.5	4.0	911	1 DLG1_RAT	O62696 rattus norv
14	246	3.5	2492	1 ATRX_HUMAN	P46100 homo sapien
15	238.5	3.4	742	1 NEB1_HUMAN	O9u1j8 homo sapien
16	238.5	3.4	2805	1 NEB1_HUMAN	P78359 homo sapien
17	237.5	3.4	3924	1 ANK2_HUMAN	Q01484 homo sapien
18	233	3.3	1411	1 TCOF_HUMAN	Q13428 homo sapien
19	229	3.3	1462	1 NKCR_HUMAN	P30414 homo sapien
20	228	3.2	2464	1 MAPB_MOUSE	P14873 mus musculu
21	227.5	3.2	960	1 DLG1_DROME	P31007 drosophila
22	221.5	3.1	771	1 CALD_CHICK	P12957 gallus gall
23	221.5	3.1	2774	1 MAPA_RAT	P34926 rattus norv
24	217	3.1	2426	1 SON_HUMAN	P18583 homo sapien
25	214.5	3.0	1190	1 ZO2_HUMAN	O9u4y2 homo sapien
26	213	3.0	1167	1 ZO2_MOUSE	O9z0h1 mus musculu
27	212	3.0	2004	1 MOZ_HUMAN	O92194 homo sapien
28	210.5	3.0	1174	1 ZO2_CANFA	O95168 canis famill
29	210.5	3.0	2459	1 MAPB_RAT	P15205 rattus norv
30	209.5	3.0	1189	1 YJH6_YEAST	P47035 saccharomyc
31	209.5	3.0	1898	1 TRHY_HUMAN	O07283 homo sapien
32	208.5	3.0	817	1 NEB2_RAT	O35274 rattus norv
33	206	2.9	2845	1 APC_MOUSE	O61315 mus musculu

34	204	2.9	1021	1 MAPA_MOUSE	O9qyrf mus musculu
35	203.5	2.9	1233	1 YF16_YEAST	P43597 saccharomyc
36	203.5	2.9	2843	1 APC_HUMAN	P25054 homo sapien
37	203	2.9	1453	1 NKCR_MOUSE	P30415 mus musculu
38	203	2.9	1654	1 PCRB_HUMAN	O94913 homo sapien
39	203	2.9	1781	1 AKAC_HUMAN	O02952 homo sapien
40	202	2.9	3969	1 HRX_HUMAN	O03164 homo sapien
41	201	2.9	2476	1 ATRX_MOUSE	O61687 mus musculu
42	199.5	2.8	1023	1 GLT_DROME	P33438 drosophila
43	199.5	2.8	1531	1 NPFS_HUMAN	O94916 homo sapien
44	197.5	2.8	1744	1 TANA_XENLA	O01550 xenopus lae
45	196.5	2.8	1359	1 ATRX_CAEEL	O9u7e0 caenorhabd

## ALIGNMENTS

RESULT 1  
PTND\_HUMAN STANDARD; PRT; 2485 AA.  
AC Q12923; Q15263; Q16826; Q15264; Q15265; Q15159;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Protein tyrosine phosphatase, non-receptor type 13 (EC 3.1.3.48)  
DE (Protein-tyrosine phosphatase 1E) (PTP-EL) (PTP-BAS) (Protein-tyrosine  
DE phosphatase PTP1L) (Fas-associated protein-tyrosine phosphatase 1)  
DE (FAP-1).  
GN PTPN13 OR PTP1E OR PTP1L OR PNP1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Breast carcinoma;  
RX MEDLINE=94350988; PubMed=8071359;  
RA Banville D., Ahmad S., Stocco R., Shen S.-H.;  
RT "A novel protein-tyrosine phosphatase with homology to both the  
RT cytoskeletal proteins of the band 4.1 family and junction-associated  
RT guanylate kinases.";  
RL J. Biol. Chem. 269:22320-22327(1994).  
RN [2]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE-Leukemia;  
RX MEDLINE=94116679; PubMed=8287977;  
RA Maekawa K., Imagawa N., Nagamatsu M., Harada S.;  
RT "Molecular cloning of a novel protein-tyrosine phosphatase containing  
RT a membrane-binding domain and GIGF repeats.";  
RL FEBS Lett. 337:200-206(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Fibroblast;  
RX MEDLINE=95014139; PubMed=7929060;  
RA Sars J., Claesson-Welsh L., Heldin C.-H., Guez L.J.;  
RT "Cloning and characterization of PTP1L, a protein tyrosine phosphatase  
RT with similarities to cytoskeletal-associated proteins.";  
RL J. Biol. Chem. 269:24082-24089(1994).  
RN [4]  
RP SEQUENCE OF 1216-2490 FROM N.A.  
RC TISSUE-Pancreas;  
RA Wang H.Y.;  
RN Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
RP STRUCTURE BY NMR OF 1361-1456.  
RX MEDLINE=20170882; PubMed=10704206;  
RA Kozlov G., Gehring K., Ekfel I.;  
RT "Solution structure of the PDZ domain from human phosphatase hPTP1E  
RT and its interactions with C-terminal peptides from the Fas  
RT receptor.";  
RL Biochemistry 39:2572-2580(2000).  
CC -I- FUNCTION: BINDS TO A NEGATIVE REGULATORY DOMAIN IN FAS THAT  
INHIBITS FAS-INDUCED APOPTOSIS.

CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: PRESENT IN MOST TISSUES WITH THE EXCEPTION OF  
 CC THE LIVER AND SKELETAL MUSCLE. MOST ABUNDANT IN LUNG, KIDNEY AND  
 CC FETAL BRAIN.  
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 5 PDZ/DHR DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
 CC TYROSINE PHOSPHATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U12128; AAB60339.1; -  
 DR EMBL; D21209; BAA04750.1; -  
 DR EMBL; D21210; BAA04751.1; -  
 DR EMBL; D21211; BAA04752.1; -  
 DR EMBL; X80289; CAA56563.1; -  
 DR EMBL; X79676; CAA56124.1; -  
 DR PDB; 3PDZ; 17-MAR-00.  
 DR MIM; 600267; -  
 DR InterPro: IPR000299; Band\_4.1.  
 DR InterPro: IPR001478; PDZ.  
 DR InterPro: IPR000387; Tyr\_phosphatase.  
 DR InterPro: IPR000242; Tyr\_prot\_phosphatase.  
 DR Pfam; PF00373; Band\_4.1; 1.  
 DR Pfam; PF00595; PDZ; 5.  
 DR Pfam; PF0102; Y\_phosphatase; 1.  
 DR PRINTS; PR00935; BAND4.1.  
 DR PRINTS; PR00700; PRTPHPTASE.  
 DR SMART; SM00295; B4; 1.  
 DR SMART; SM00288; PDZ; 5.  
 DR SMART; SM00194; PTPc; 1.  
 DR PROSITE; PS00660; BAND\_4.1; FALSE\_NEG.  
 DR PROSITE; PS00661; BAND\_4.1.2; FALSE\_NEG.  
 DR PROSITE; PS00057; BAND\_4.1.3; 1.  
 DR PROSITE; PS00106; PDZ; 5.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; FALSE\_NEG.  
 DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 1.  
 DR Structural protein: Cytoskeleton: Hydrolase; Repeat; 3d-structure;  
 KW Alternative splicing; Coiled coil.  
 FT DOMAIN 56 59  
 FT DOMAIN 585 879  
 FT DOMAIN 2237 2485  
 FT DOMAIN 379 399  
 FT DOMAIN 469 504  
 FT DOMAIN 1775 1804  
 FT DOMAIN 2057 2085  
 FT DOMAIN 1093 1178  
 FT DOMAIN 1368 1452  
 FT DOMAIN 1501 1588  
 FT DOMAIN 1788 1868  
 FT DOMAIN 1882 1965  
 FT DOMAIN 1742 1749  
 FT ACT\_SITE 2408 2408  
 FT VARSPIC 884 1074  
 FT VARSPIC 1056 1074  
 FT CONFLICT 1134 1135  
 FT CONFLICT 1216 1229  
 FT CONFLICT 1238 1239  
 FT CONFLICT 1357 1357  
 FT CONFLICT 1362 1363  
 FT CONFLICT 1383 1383

FT CONFLICT 1538 1538 P -> A (IN REF. 3).  
 FT CONFLICT 1649 1649 R -> K (IN REF. 4).  
 FT CONFLICT 1698 1714 KSDPTCTMYYPOKI -> RVAKIPVCPPIIKR  
 FT CONFLICT 1797 1797 (IN REF. 4).  
 FT CONFLICT 1856 1857 G -> A (IN REF. 3).  
 FT CONFLICT 2069 2069 AA -> G (IN REF. 4).  
 FT CONFLICT 2206 2210 A -> S (IN REF. 4).  
 FT CONFLICT 2485 AA; 276903 MW; 8DIB31597C6962B CRC64;  
 SQ SEQUENCE

Query Match Best Local Similarity 4.68; Score 324.5; DB 1; Length 2485;  
 Matches 193; Conservative 134; Mismatches 361; Indels 245; Gaps 32;

QY 80 DEODPHHGDDGNSASTGTOSPEIFGSLG-----TNNVAFQPYQATSEI--E 126  
 DB DAQDI-----EASFRSLNLQAESYGFNMGRALSTGSLASTINKLAVRPLVQAEILKR 934  
 QY 127 VTPSVLRANMPLH--VRRSDPAL-----IGLSTVSDFSSESESRKPKPTKMTAGF 179  
 DB 935 LSCSELALYDPLQNSRKEDKASWEKPRMKSYSYHDLQASLYHRKN----- 984  
 QY 180 LKONTAGSPKTCDRKKNENTRSLPR-DTSWMSQFORDNARSSLASHPWGKMLEKQEQ 238  
 DB 985 VIVMPEPPQVVALVCKPESHMSRSDAESLAGVTYKLNSKSVASLIN-----RSP 1034  
 QY 239 DEDCTEDNSRVEPVGHA-----DTGLEHIFNFSLDLV-----YKIV 275  
 DB 1035 ERRKHEDSSSIDPQAYVLGNTMHSNGSSSQVYLKENDVLRKMSIVSPEREITLY 1094  
 QY 276 EVPMDG-PGLIIVPVSANGRTLLYLKLEGGKAEHMLFENDCIYRINDGILRN 334  
 DB 1095 NLKKDAKYGIGFQIGEEKMGRILGIFISVAPGPADLDCLKPGRLISVNSVSLBG 1154  
 QY 335 RREDOAHMFRQA-----MRPIIMFHV-----VPAA 361  
 DB 1155 VSHHAALIELONAPEDVTLVISQPKKIKVSTVYVHLNENKNYKSSYMODSAIDSS 1214  
 QY 362 NKROY-----EQLSSEKNYYSRPF----- 382  
 DB 1215 SKDHMSRQTLNHIENSPGSGGLREGSLSSQDSRTESASLSQGVNGFASHLIGQTW 1274  
 QY 383 -----SPDSQYIDNRSVNSAGLHTVQ---RAPRLNHPEDID-SHSLPISAHSPSKP 432  
 DB 1275 QRSQHSPPSPVYSKATKEKFTPDNOSKTKKPGISDVTYSDSGDDMDQEAATYSSQDH 1334  
 QY 433 SAPASAPQVVFSTYVSSGYN-----KIKGRNLIOLEKTEGIGFESIT---SHDVTIGG 484  
 DB 1335 QTPKQSSSVNTSKMNFKFTSSPPKPGDIFVEELAKNDNSLGISVGTGVMTSVRHGG 1394  
 QY 485 SAPIVYKNILPRGCAIODRLKAGDRILEVNGVLYGKSOEYVSLSRKMGTVSLV 544  
 DB 1395 ---IYKAVIPQGAESDGRHKKDRVLAVNGVSLSEATIKQAEVTLRNT---GQVYVHL 1448  
 QY 545 FROEDAIFRPE-----LMAEPSOMQIP---KETKAEDDIYLPDGTREPLFEV 591  
 DB 1449 LEKQOSPTSKHEPVTPQCTLSDQNAOGQPEKVKTYQKDYKDSV-----TEENTFEV 1502  
 QY 592 PL-NDSSAGLGVSVKGNRSKENHADLGIFVKSTINGGAASKGRILRVNDQLAVNGESL 650  
 DB 1503 KLFKNSGLGSPSRENLLPEQINMSIYVKKLPGQPAESGKIDVDVILKVGASL 1562  
 QY 651 LKKNQAMETLRRSMSTEGNKGMIQLIVARRISCKNELSPGSPPEGLPIETALDDR 710  
 DB 1563 KGLSQQEVISALGTAPAE-----VFLLCR-----PPGVLPETIDAL--- 1600  
 QY 711 ERRISHLYGIGIGLDESBSRNALSR-----MGESKYQL-SPTVNMPODDTV 759  
 DB 1601 -----LPLQSPAVLYLNNSSKDSQPSVCVQSTSDENEMSKKCKCKSPSRSDYSOSS 1656  
 QY 760 ILEDRLPYLPHILSDQS-SSSHDVGCVFTADAGTWAKAALS----- 801

DB 1657 GSGEDLTATAPANINSNSTWSSALHOTLSNMVSOASHHEAPKSOEDTICTMFTYPOKIPN 1716

QY 802 -----DSADCSLSPVDVPLAFQREGFGROSMS 829

DB 1717 KPEFDSNNSPLPDPMAPOQSQYPOSESASSSS 1749

RESULT 2

AF6\_HUMAN STANDARD; PRT; 1816 AA.

AC P55196; 075087; 075088; 075089; Q9NU92;

DT 01-OCT-1996 (Rel. 34, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE AF-6 protein.

GN MLT4 OR AF6.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RX MEDLINE=94061833; PubMed=8242616;

RA Pressed R., Gu Y., Alder H., Nakamura T., Canaan O., Saito H.,

RA Huebner K., Gale R.P., Nowell P.C., Kuriyama K., Miyazaki Y.,

RA Croce C.M., Canaan E.;

RT "Cloning of the ALL-1 fusion partner, the AF-6 gene, involved in

RT acute myeloid leukemias with the t(6;11) chromosome translocation.";

RL Cancer Res. 53:5624-5628(1993).

RN [2]

RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RC TISSUE=Fetal brain;

RX MEDLINE=98344142; PubMed=9679199;

RA Saito S., Matsushima M., Shitahama S., Minaguchi T., Kanamori Y.,

RA Minami M., Nakamura Y.;

RT "Complete genomic structure, DNA polymorphisms, and alternative

RT splicing of the human AF-6 gene.";

RL DNA Res. 5:115-120(1998).

RN [3]

RP SEQUENCE OF 337-1816 FROM N.A. (ISOFORM 2).

RA Williams S.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

CC FUNCTION: MAY ACT AS AN INTRACELLULAR SIGNALING COMPONENT

CC CONTROLLED BY RAS SIGNALING PATHWAYS.

CC -1- SUBUNIT: BINDS DIRECTLY TO ZO-1 AND OCCUDIN.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1, 2 (SHOWN HERE) AND 3; ARE

CC PRODUCED BY ALTERNATIVE SPLICING.

CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL

CC TRANSLOCATION T(6;11)(Q27;Q23) THAT INVOLVES MLT4 AND MLL/HRX.

CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.

CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AB011399; BAA32484.1; -

DR EMBL; AB011399; BAA32484.1; -

DR EMBL; AB011399; BAA32485.1; -

DR EMBL; U02478; AAC50059.1; -

DR EMBL; AL049698; CAB76850.1; -

DR HSSP; Q12923; 3PDZ.

DR MIM; 159559; -

DR InterPro: IPR002710; DIL.

DR InterPro: IPR000253; FHA\_domain.

DR InterPro: IPR001478; PDZ.

DR InterPro: IPR00159; RA.

DR Pfam: PF01843; DIL; 1.

DR Pfam: PF00498; FHA; 1.

DR Pfam: PF00595; PDZ; 1.

DR Pfam: PF00788; RA; 2.

DR ProDom: PD003376; DIL; 1.

DR SMART; SM00240; FHA; 1.

DR SMART; SM00228; PDZ; 1.

DR PROSITE; PS0106; PDZ; 1.

KW Chromosomal translocation; Proto-oncogene; Alternative splicing.

FT DOMAIN 36 206

FT FT 425 491

FT DOMAIN 804 910

FT FT 991 1077

FT DOMAIN 162 174

FT FT 1349 1356

FT DOMAIN 1371 1376

FT FT 1561 1571

FT SITE 26 26

FT FT 1588 1611

FT FT 1612 1816

FT FT 1666 1743

FT FT 1744 1816

FT FT 373 373

FT FT 391 391

FT FT 744 744

FT FT 1031 1031

FT FT 1408 1408

SO SEQUENCE 1816 AA; 205604 MW; EB1FE7F04879CEBF CRC64;

Query Match 4.2%; Score 296; DB 1; Length 1816;

Best local similarity 20.1%; Pred. No. 9.8e-07;

Matches 340; Conservative 217; Mismatches 536; Indels 600; Gaps 86;

QY 49 HLEHEDGGLIDLDLDCV-----ADKRLVAVFDEDDPHI----- 86

DB 307 HSDKAKAKETI-LDDDECPQLIFREWPSDKGLVFLKRPDHPKTKKHLEGKTPKG 365

QY 87 -GGDGSAST-----GTQS---PEIFG-----SELGT-----NNVSAF 116

DB 366 KERADSGYGSTLPPKELPYVELSPDGSQSDKPKLYRLQLSVTEGTEKLDNLSIQLE 425

QY 117 ---QPYQA-----TSEIETVPSVLNANMPLHYRASSDPALI--GLTSVSDN-FSSSE 164

DB 426 GPEIGHHCDLTMDGVVYITPRSMDEIFYVEGQRISETTMLOSGMKVQGCASHVKEFVD 485

QY 165 PSKKNP-TRKSTTAGFLKONTAGSP-----KTCRRKD-ENTYSLPRDTSNMSNOFORN 217

DB 486 PSDNDHAKRSYDVGMLVNGPRHKPGIIVQETTPDLGDGHSHTALP--TSKSTRLDSR 543

QY 218 ARSLASAH-----PMVG-----KMLEKQEDDEGTE-----EDNSRVEP 252

DB 544 VSSASTAERGKVKPIRVEQOPDYRRQESRTDASGPELILPASIEPRESSDSLSAI 603

QY 253 VGHADTGLEHI---PNFSLDMMVKLEVPRDGGPLGIHVFPFSGRGRTLGLLVKRLKNG 309

DB 604 INTNSGVHFKLSPIYVILMACRYV-LSNQYRP---DISP-TERHNKIVAVVKNKV--- 655

QY 310 GKAEHNLRENDICV-----RINDGDLNRKFEQAQHMFRQAMKTPITWIVPAANK 363

DB 656 --SMMEGVIOKOKINAGALAFMANNASELIN-FIKODRDLISRTTLAODVLAHLYOMARK 712

QY 364 EGYEOLSOSEKKNYSSRSPDSQYIDNRSVNSAG-----LHTVQ-----RARLNH 410

DB 713 YLVHCL-QSELNMYMPA-----FLDDPEENSLQRPKIDVDVLTLLTGAMSLILRCRVN- 763

[illegible]

```

RESULT      3
DLG3_MOUSE
STANDARD:
PRT:      849 AA.

```

large homolog 5).  
DE  
CN  
PLG3 OR DLGH3.  
Fute]eastomi;

RC STRAIN=C57BL/6; TISSUE=Brain;  
Makino S., Yagi T.;  
GenBank/DBJ databases.

-1- FUNCTION: INLEADS  
CC CONTAINS 3 PDZ/DHR DOMAINS  
CC SUBUNIT NR2B.

CC - SIMILARITY: CONTAINS A GUANINATE ENO  
CC - SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEIN  
CC - SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEIN

CC this SWISS-PROT entry is available on its own without restrictions on its use.

CC between the Swiss Institute of Bioinformatics Institute. There are no restrictions on its use.

CC between the Swiss Institute of Bioinformatics Institute. There are no restrictions on its use.

modified and this license agreement (cc entities requires a license@isb-sib.ch)

CC BAA13249.1; -  
DR D87117; EMBL; D31016; 1BE9.

DR InterPro; IPR000615; SH3.  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR001473; SH3.

DR	PFam;	PDZ;	3.
DR	PF00595;	SH3;	1.
DR	PF00018;	SH3;	1.

DR	SMART;	SM00228;	SH3; 1.
DR	SMART;	SM00326;	CHYNIATE KINASE-1; 1
DR	SMART;		

DR	PROSITE: PS50106; PDZ; 3.
DR	PROSITE: PS50002; SH3; 1.

FT	FTT	DOMAIN	PDZ
149	222	244	2.
330	484	244	3.

FT	849	GUANI	EF3EF2D7513538EE	CRC64
DOMAIN	659			
DOMAIN	849 AA:	93482 MM:		
FT				

Query	Match
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	1
30	1
31	1
32	1
33	1
34	1
35	1
36	1
37	1
38	1
39	1
40	1
41	1
42	1
43	1
44	1
45	1
46	1
47	1
48	1
49	1
50	1
51	1
52	1
53	1
54	1
55	1
56	1
57	1
58	1
59	1
60	1
61	1
62	1
63	1
64	1
65	1
66	1
67	1
68	1
69	1
70	1
71	1
72	1
73	1
74	1
75	1
76	1
77	1
78	1
79	1
80	1
81	1
82	1
83	1
84	1
85	1
86	1
87	1
88	1
89	1
90	1
91	1
92	1
93	1
94	1
95	1
96	1
97	1
98	1
99	1
100	1

4.28; Score 294; DB 1; Length 849

Best Local Similarity 20.8%; Pred. No. 4.5e-07;  
Matches 198; Conservative 130; Mismatches 314; Indels 310; Gaps 40;

```

QY 222 LSASHPMVGKLEKQEDQEDGTEEDNSHVEVGHADTGLEHHPNPSL---DDMYLVLEVP 278
Db 98 LNSGSP--GMMPECTCTRDWYEQ--ASPAFLVNPALF--PILSTVSGSGMFEYELV 151
QY 279 NDGGPLGIHVVPFSARG-----RTGLLVKLEKGGKAKHENLFRENDCIYRINDG 331
Db 152 LERNNSGL--GFSTAGCIDNPHVDDPGIFITKLIIPGCAAMQRLGVNCCVLAVNEVD 208
QY 332 LRNRFEQAQMFQAMPTPIIMFHVPAANKQEYEQLSQSEKNYYSRSRSPDQYIDN 391
Db 209 V-----AGLHTVQARAPRLNHPPEQIDSHSLPHSAHPGKPPAPASAPONVSTYVS 215
QY 392 RSVNS--AGLHTVQARAPRLNHPPEQIDSHSLPHSAHPGKPPAPASAPONVSTYVS 448
Db 216 RAVELAKLAG--PYVRLVYRRQPP-----PETI-----242
QY 449 SGYNTKKIGKRLNIQKKGTEGLGFSITS--RDVTIGGSAPLYYKNILPRCAIADGRLK 506
Db 243 -----MEVNLTKGPKGLGSIAGIGNQHIIPGDNSTIYTKIIEGGAARQDRLQ 291
QY 507 AGDRLEIYNGVDLYGKSGEEVYSLLRSTK-----MEGVSYSLVFEQEDAFHRE--- 555
Db 292 IGDRLAVNNNTLQDVRHEEVASLKNSTDWYLYKVAKPGSIHL---NDMYAPPDYAS 346
QY 556 -----LNAEPSQM-----QIPKETEKADEEDIVLTPDGT 583
Db 347 TETLADNHHSNLSGLGAVESKVTYPPAPQVPTPTKSPTRIMLAE-----DFT 399
QY 584 REPLTEVP---LNDGSGAGLGAVYKGNRSKENHADLGIYFKSIINGAASKRGLRYND 640
Db 400 RE-----PRKTIHKGTGLGFINVGEGDE---GIFVSFIIAGPADLSGLRQD 448
QY 641 OLIVNGESSLIGKTNQDMETLRRSMSTEGNKRGMIOIIVARIKCMELKSPGSPGPE 700
Db 449 RILSYNGVNLNNAHEQAALAKRA-----GQSVTVAVO-----YRPEXY 488
QY 701 LPIETALDD--RERRISHSLYSGIEGLDESPNNALSRIMSGSGYQLSPVNMPPQDPTV 759
Db 489 SRFEKIHDLAEQAMNSMSSGSLRTSEKSLYVRLF-----DYD 531
QY 760 IIEEDRLFPVLPPLHSDQSSSSSHDV--GFVTADACTWAKAAL---SDADCSLSPDVP 814
Db 532 RTDSCLP-----SQGLSFSGDILHVINASDDEMMQARLVTPHGSEQIGVIP--- 580
QY 815 VLAPQREGFGRQSMSEKRTKQPSDASQLDYFKTRKSKMDLGIADETKILNTYDDOKAGSP 874
Db 581 -----SKKRYEK--KERARLKYVFHARTGQIESNRDPFGIS--DDYVGAKN 623
QY 875 SRDVGPSLGLKKSSLSLESLQTAFAVEVTLNGDIPFHRPRPRIIRG---RQCNEFPRAIDK 931
Db 624 LKGYTSMNSDSESS--KQEDALILSEPYTKQEIHYARPVILIGPMKDRVNDLISEFPH 682
QY 932 SY-----DKPAVDDDEGM-----ETLEEDTESS--RSGR-----ESVS 964
Db 683 KFGSCVPTHTPRRDNEVDGQDYHFVVSREQMEKDIDMKFEAGOFNDNLGTSTISQVR 742
QY 965 TASDQPSHSLEKQNMNGKQKTKTRKKDKYKKEKKDKRDKKAKKAGKMLKGLGDMF 1024
Db 743 AVAEGKACIT--LDVSGNAIK-----RLQQAOLYPIAIFIKPKSIEALMEKN 787
QY 1025 RFGKRRKDKI--EKTGKIKIQESFTESEERLIKMOEOERIQATREFREROA 1075
Db 788 RROTEYQANKIFDKA--MKLEGEFGEYFATLYQSGDSLEETIKIKIILEDQS 837

```

RESULT 4  
DLG4\_HUMAN STANDARD: PRT; 767 AA.  
ID DLG4\_HUMAN AC P78352: Q92941: Q9UKR8:  
01-NOV-1997 (Rel. 35, Created)

```

DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Presynaptic density protein 95 (PSD-95) (Discs, large homolog 4).
GN DLG4 OR PSD95.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=97432822; PubMed=9286702;
RA Stachakis D.G., Hoover K.B., You Z., Bryant P.J.;
RT "Human postsynaptic density-95 (PSD95): location of the gene (DLG4)
RT and possible function in nonneural as well as in neural tissues.";
RL Genomics 44:71-82(1997).
RN [2]
RP REVISIONS.
RC TISSUE=Mammary gland;
RA Stachakis D.G., Hoover K.B., You Z., Bryant P.J.;
RT Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20047407; PubMed=10582582;
RA Stachakis D.G., Udar N., Sandgren O., Andreasson S., Bryant P.J.,
RA Small K., Forsman-Semb K.;
RT "Genomic organization of human DLG4, the gene encoding postsynaptic
RT density 95.";
RL J. Neurochem. 73:2250-2265(1999).
RN [4]
RP SEQUENCE OF 81-401 FROM N.A.
RC TISSUE=Brain;
RA Breuman J.E., Bredt D.S., Parkinson J.F., Manzana W.P., McClary J.A.;
RT Submitted (AUG-1996) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
CC - SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY
CC ON THE PRESYNAPTIC SIDE (WAS ORIGINALLY THOUGHT TO BE
CC POSTSYNAPTIC).
CC - TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.
CC - SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC - SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC - SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC - SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC EMBL: U83192; AAC52113.1; -.
CC EMBL: AF156495; AAD56173.1; -.
CC EMBL: U68138; AB07736.1; -.
CC HSSP: P31016; 1BE9.
CC MIM: 602887; -.
CC InterPro: IPR000619; Guanylate_kin.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001452; SH3.
CC Pfam: PF00625; Guanylate_kin; 1.
CC Pfam: PF00595; PDZ; 3.
CC Pfam: PF00018; SH3; 1.
CC SMART: SM00072; Gukc; 1.
CC SMART: SM00228; PDZ; 3.
CC SMART: SM00326; SH3; 1.
CC PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
CC PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
CC PROSITE: PS50106; PDZ; 3.
CC PROSITE: PS50002; SH3; 1.
CC SH3 domain; Repeat.
CC DOMAIN 108 194 PDZ 1.

```





```

Db 210 -----SEVYHSAVBALEAG-PVYRLVVRRRQP----- 238
OY 431 PPSAPASAPONFESTVSSGYMTKKIGKRLNIOLKKGTEGFSITS--RDVTIGGSAPI 488
Db 239 -----PETI-----MEYNLKGPRGLGFSINGIGNHIDPDSNI 273
OY 489 YKKNILPFGCAIQDGLKAGDRLEIYGVLDYKGSOEYVSLRSTKMGVSLVLRQE 548
Db 274 YTKIIEGGAOKDRLOIGDRLLAVNNTNLODVREHEAVASLTKNT--SDMYLKVAPR 330
OY 549 DAFHRELAERPSQK-----IRK 567
Db 331 GSLILNDMAIPDYASTFTALADNHSNLSGLYGAVESKYTPAPPOVPTRPSPIR 390
OY 568 ETKAEDEDIVLPDGTREFLFEVP---LNDGSAGLGYGVKGNRKENHADLGIKFSI 624
Db 391 IHLAEE-----DFTRE-----PKIILHGSGTGLGNTIYGEGE-----GIVVSTI 432
OY 625 INGAASKDGLRLVNDQILVANGESILGKTNDAMETLRRSMSTEGNKKRMIOIIVARI 684
Db 433 IAGGPAIDSGELRGDRILSTVNGVNLRNATHEQAAALAKRA-----GQSTVIYAQ-- 482
OY 685 SKCNELKSGSPGPELPLETALDO-REKRISHSLYSGIEGLDESPRMAALSRIKESG 743
Db 483 -----YRPEYSRPESKIHDIREQMNSMSGSGSLTSEKRLIYVAALF----- 528
OY 744 KYQLSPVPMPODDVYIIEDRLPVLPPLHSDQSSSSSSHDV-GFYTADAGTWAKAI-- 800
Db 529 -----DYDRTRDSCLP-----SQGLSFSGDILHYNADDEWQARLYT 568
OY 801 --SDSADCSLSPDVPVLAIFQREGRGROSGMSERKTRKQFSDASOLDVYKTRKSKMDLGA 858
Db 569 PHGESQIQIVP-----SKRKVEK-KERARKLVKFKHARTGMIESNR 609
OY 859 DETKLTNTVDQAGSPSRDVGPSLGLKSSLSLEJOTNAEVLINDIFHRRPPIIING 918
Db 610 DPEGLS--DDYGAKNLKVNTSNTSDSESS--KGOADALISTEPTVROGIRHARPIIIG 666
OY 919 ---RCGNESFRAIDKSY-----DKPAVDDDEGEM-----ETLEEDTEESS--RS 958
Db 667 PKMDRVNDOLISEFPHKFCSCVPHTTRPRRDNEVQODHFVYSRQMKDIDQNKFIET 726
OY 958 GR-----ESVSTASDOPSHSLEROMNGOEKDKTRKKDKTGKKKDRDEKO 1008
Db 727 GOFNDNLVSTSIQSVAAVERGKHCI-LDVSGNAIK-----RLQQAOLYPI 771
OY 1009 KKKAKGMKGLGDMPREFGKHRRDKI-EKTKIKIQESFTSEERIRKQDERIQART 1067
Db 772 AIFIKPKSTIEALMENKROTYPQANKIFDKA--MKLEDFEGEFTYALVQDLSLETYKRI 829
OY 1068 REFERRQA 1075
Db 830 KQIEDQS 837

```

```

RX MEDLINE-98031926; PubMed-9362513;
RA Nakanishi H., Obatahi H., Satoh A., Wada M., Mandai K., Satoh K.,
RA Nishio H., Matsura Y., Mizoguchi A., Takai Y.;
RT "Neurabin: a novel neural tissue-specific actin filament-binding
RT protein involved in neurite formation.";
RL J. Cell Biol. 139:951-961(1997).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE-Brain;
RX MEDLINE-98318661; PubMed-9653190;
RA Burnett P.E., Blackshaw S., Lai M.M., Qureshi I.A., Burnett A.F.,
RA Sabatini D.M., Snyder S.H.;
RT "Neurabin is a synaptic protein linking p70 S6 kinase and the neuronal
RT cytoskeleton.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8351-8356(1998).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE-Brain;
RX MEDLINE-20054471; PubMed-10585469;
RA Macmillan L.B., Bass M.A., Cheng N., Howard E.F., Tamura M.,
RA Strack S., Madzinski B.E., Colbran R.J.;
RT "Brain actin-associated protein phosphatase 1 holoenzymes containing
RT spinophilin, neurabin, and selected catalytic subunit isoforms.";
RL J. Biol. Chem. 274:35845-35854(1999).
RN [4]
RP INTERACTION WITH TGN38.
RX MEDLINE-99445568; PubMed-10514494;
RA Stephens D.J., Bantling G.;
RT "Direct interaction of the trans-Golgi network membrane protein,
RT TGN38, with the F-actin binding protein, neurabin.";
RL J. Biol. Chem. 274:30080-30086(1999).
RN [5]
RP INTERACTION WITH PPL, PHOSPHORYLATION SITE SER-431, AND MUTAGENESIS.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE-99435770; PubMed-10504266;
RA McAvoy T., Allen P.B., Obatahi H., Nakanishi H., Takai Y.,
RA Greengard P., Nairn A.C., Hemmings H.C. Jr.;
RT "Regulation of neurabin I interaction with protein phosphatase 1 by
RT phosphorylation.";
RL Biochemistry 38:12943-12949(1999).
CC -1- FUNCTION: BINDS TO ACTIN FILAMENTS (F-ACTIN) AND SHOWS CROSS-
CC LINKING ACTIVITY. BINDS ALONG THE SIDES OF THE F-ACTIN. MAY BE
CC INVOLVED IN NEURITE FORMATION. INHIBITS PROTEIN PHOSPHATASE 1-
CC ALPHA ACTIVITY. MAY PLAY AN IMPORTANT ROLE IN LINKING THE ACTIN
CC CYTOSKELETON TO THE PLASMA MEMBRANE AT THE SYNAPTIC JUNCTION.
CC -1- SUBUNIT: POSSIBLY EXISTS AS AN HOMODIMER, HOMOTRIMER OR AN
CC HOMOTETRAMER. INTERACTS WITH F-ACTIN, PROTEIN PHOSPHATASE 1 (PPL),
CC NEURABIN-II, TGN38 AND P70-S6K.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ALSO FOUND IN THE SOLUBLE
CC SYNAPTOTOMAL FRACTIONS.
CC -1- TISSUE SPECIFICITY: BRAIN, AND WIDELY EXPRESSED IN NEURAL TISSUE.
CC HIGHLY CONCENTRATED IN SYNAPSES OF DEVELOPED NEURONS. IN
CC DEVELOPING NEURONS, CONCENTRATED IN THE LAMELLIPODIA OF THE GROWTH
CC CONE.
CC -1- DOMAIN: INTERACTS WITH P70-S6K VIA ITS PDZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DRH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U72994; AAC53454.1; -.
CC HSSP: Q12923; 3PDZ.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001660; SAM.
CC Pfam: PF00595; PDZ; 1.
CC Pfam: PF00536; SAM; 1.
CC SMART: SM00228; PDZ; 1.
DR

```

DR SMART: SM00454; SAM; 1.  
 DR PROSITE: PS0106; PDZ; 1.  
 KW Neurogenesis; Actin-binding; Cytoskeleton; Coiled coil;  
 FT Phosphorylation.  
 FT DOMAIN 1 144 ACTIN-BINDING.  
 FT DOMAIN 505 593 PDZ.  
 FT DOMAIN 989 1052 SAM.  
 FT DOMAIN 598 628 COILED COIL (POTENTIAL).  
 FT DOMAIN 674 825 COILED COIL (POTENTIAL).  
 FT DOMAIN 1036 1091 COILED COIL (POTENTIAL).  
 FT DOMAIN 426 503 INTERACTS WITH PROTEIN PHOSPHATASE 1.  
 FT MOD\_RES 598 1091 INTERACTS WITH TGN38.  
 FT MOD\_RES 461 461 PHOSPHORYLATION (BY PKA).  
 FT MUTAGEN 461 461 S>E: 35-FOLD DECREASE IN INHIBITION OF  
 FT MUTAGEN 461 461 P1-ALPHA.  
 FT MUTAGEN 516 517 GI->AA: ABOLISHES P70-S6K BINDING.  
 FT SEQUENCE 1095 AA: 122735 MW: 58E3650320B3D61 CRC64;

Query Match Best Local Similarity 4.2%; Score 293; DB 1; Length 1095;  
 Matches 269; Conservative 163; Mismatches 477; Indels 470; Gaps 56;

75 LVAVFEDDDHHGGDSTASSTGTGSPETFGSELTN-----NY-----SAFQPYATSEIE 126  
 30 LKSTPKKPP-----DEQKTEGEGSQSQRKYGSNVNRKLNLPQMGMPENNAITIA 85  
 127 VT-----PSVLRAAM-PLHVRSSDPALIGLSTYSV--SNFSS--EPPS-----RKN 169  
 86 KTKGKRPSSPQKRMKPKFEVEKTDGSGVYKLESSVSEKISFDMHGPYAKFTETRMK 145  
 170 PTWSTTAFLKONTAGSPKTCDDKNDENYRSLPRD-----TSMNSQFORARSLA 224  
 146 PERSHESG-----QNNRHS-----KKEKAGAEFQDEMGSSKRRSSDLSLGRTEA 197  
 225 SHPNTGKMLEKQEDDED-----GTEEDNSRVEPYGHADTLEIPIFSLDDMYKLVYFN 279  
 138 VSPVTSQLSAVFENSESGAITTPGKANSVNYVTGHPPLNLPVTVNTND----- 247  
 280 DGGPLGIHVFPFARGRTGLLVKLEKGAHEHNLFRENDCIYRIDNDLRMRREQ 339  
 248 -----TFG-----RLKDSNRSRSSKKQ 264  
 340 A-----QMFQAMRTPIIMFHVPAANKEQYBOLSOSEKNNYSRSPDSQYIDNRSVN 395  
 265 ATDTEPEKSEAVPEVAOKGTSLSLPSEROLSTEADVTAPDPTDSTDKDGP 324  
 396 SAGLHTVORAPRLNHPPEID-----SHSLRPHSAHPSG-----KPPSAPASA 438  
 325 SAESQAMPKSNLTLSRKEPLEDEANVVSSEAEQFORADLTGGDDLSPDASASCKEY 384  
 439 --PQNFEST--VSSGYNTKTKIGRLNIOL-KKGTBGLSPSTISRYVTIGSAPLYKN 492  
 385 PEDSNFESHSYHMSDYNYRVRSRNSDMGEGTE----- 421  
 493 ILPRGATIDGRLKAGDILFVNGV-----DIVKSOBEVVSLLRSTMEGTSLVLE 545  
 422 -----QD-----EGDSDSDNNITYQPDMEYSELVGLPOEEIIPARKIKF-SCAPIKYF 468  
 546 R--QEDAHFRELAEPSQOIPRETKAEDEDIYLPDGRRELTFEYVLANSGSA--- 599  
 469 NITYSMEDYDRNDVDVVAASAEYELKEKREKLELP-----VELEKDDGIGISII 520  
 600 GLVSVKGNRSKENHADLGFVKSIIINGGAASDGRRLRVNDLIANGESLIGKTNDQAM 659  
 521 GMGVGADAGLEK-----LGTFVATVEGGAARDGRIVQNDIYEVDDISLVATQONFA 575  
 660 ETLRSTSTGKRMQILIVANRISKNEKLSQSPSPGPELPITADD--RERRISHS 717  
 576 TVLR-----NKKGVNREYGR-----EKPOVSEVAKQLSOTLEORRERRELLER 620  
 718 LYSIGIGLDESPSRNAALSRINGESGKQLSTYVAMPDDDTVIIIEDRLVLPPLHLSDS 777

Db 621 HYAAYDADD-----ETGEVATDER-----DDEVGPILP----- 649  
 QY 778 SSSSHDVGVVTLADACTWAKAA--ISDSADCSLSPDVPVLAFOREGCRGRSMSEKRTQ 835  
 Db 650 -----GGDMAIEVEFELPENEDMFSDDL-----TSKLSHK 680  
 QY 836 FSDASQIDFVTKRKSMDLGIADETKINTVDQKAAGSPSDVPSGLKSSLSLSIQ 895  
 Db 681 FKEL-QIKNAVTE-----AEIQKLT-----KLQA 704  
 QY 896 AVAEVTLNDIIFPHRPRPIINGRCNESFPAIDKSYDKPAVDDDEGNETLEEDTERS 955  
 Db 705 AE-----NEKVMLEKKNQLOQNIENKENVKL----- 733  
 QY 956 SRSRGESVSTASDPSHSLERQMNQNGDKDTPRKDKTK----- 997  
 Db 734 -----ESYWIEAQTLCHTVNEHLEKETOSYOALEKTKYKAKKLINDFOQKELDFIRQEV 788  
 QY 998 EKKKDRDEKDKMKAKKGMKGLGDM-----FR-----GKHKD 1032  
 Db 789 EKKKLEVEKKAHLVEVQGLVRIKDLEAEVFRLLKQNGTOVNNNNNIEFRSPSPETVSKG 848  
 QY 1033 DKIE-KTKIKIQESFTS-----EERIRMKQOEKIQAKTEFRRQARERDYAEI 1083  
 Db 849 DTMENVEVKQTSQDGLSDQLNEAVPETERLDSKALKTRALSVNRQRPTFR----- 903  
 QY 1084 QDFHRTGCDDELMYGVSYSBEGSMALNR-----POSPRGHMDLTVYQVKR 1134  
 Db 904 -----LYSVSTDEOSLERKNFTFNDDESPSTSSAD-LSGLGAEPTPG 949  
 QY 1135 NSKRPVDSKSTSPSNHRIQRLROEQOAKODEVEDEVRRTTYFEDQWPANRATQGR 1194  
 Db 950 LSOALSISDESL-----DWI-----DDELTD-----GQSPK 977  
 QY 1195 HVSVEVQMRQROEERESQOARQYSSLPKRSKNAASVSDSWEQ-----NYSPE 1248  
 Db 978 HTQS-----QSRVHENSVOQVSHWLVGLSLDYVEFSAQNISGQLQLDGNTLKL 1031  
 QY 1249 GPOSAKENRYSSTIGSGNGLGHPNARVMEPTQSLRQORRREQ--MKKOPPE 1305  
 Db 1032 GMTSSODRALVKKLKEKMSLE-----KARKAQEKMEKQEKLRKREDOBOMRKSSE 1086

RESULT 7  
 DLG4\_MOUSE STANDARD; PRT; 724 AA.  
 AC 062108;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Presynaptic density protein 95 (psd-95) (Presynaptic protein SAP90)  
 DE (Synapse-associated protein 90) (Disks, large homolog 4).  
 GN DLG4 OR DLG4 OR PSD95.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DBA/2; TISSUE=Brain;  
 RA Kohmura N., Yagi T.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 CC - FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR  
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS (BY SIMILARITY).  
 CC - ON THE PRESYNAPTIC SIDE (BY SIMILARITY).  
 CC - SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.  
 CC - SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC - SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: D50621; BAA09297.1; -  
 DR HSSP: P31016; 1BE9.  
 DR MGD: MGI:1277959; D1gh4.  
 DR InterPro: IPR000619; Guanylate\_kin.  
 DR InterPro: IPR001478; PDZ.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00595; PDZ; 3.  
 DR Pfam: PF00018; SH3; 1.  
 DR SMART: SM00072; Gukc; 1.  
 DR SMART: SM00228; PDZ; 3.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS00856; GUANYLATE\_KINASE\_1; 1.  
 DR PROSITE: PS50052; GUANYLATE\_KINASE\_2; 1.  
 DR PROSITE: PS50106; PDZ; 3.  
 DR PROSITE: PS50002; SH3; 1.  
 DR SH3 domain; Repeat.  
 FT DOMAIN 65 151 PDZ 1.  
 FT DOMAIN 160 246 PDZ 2.  
 FT DOMAIN 313 393 PDZ 3.  
 FT DOMAIN 428 498 SH3.  
 FT DOMAIN 534 724 GUANYLATE KINASE.  
 SO SEQUENCE 724 AA; 80472 MW; 7EFC99E1FF90BA CRC64;

Query Match 4.1%; Score 291; DB 1; Length 724;  
 Best Local Similarity 21.6%; Pred. No. 5.1e-07;  
 Matches 142; Conservative 89; Mismatches 205; Indels 220; Gaps 26;

QY 238 ODEGTEDNSVEPVGAADGLEIHPNS-----LDDMKL 274  
 DB 15 ODED-----TPLEHSPA--HLPNQNSPVIYNTDLEARGYELQVNGTEGEM 63  
 QY 275 VVPNDGPGLIHVVPFSARGRT--G-----LLVKKLEGGKAKHEMLFRENDCIYR 327  
 DB 64 EETLERNGSL--GFSLAGTDNPHIGDDPSITTKIIPGAAQAQDRLVNSILFV 120  
 QY 338 NDGDLNRFRFOQHMFROAMRTPIWPHVVPANKEQYEOLSOSEKNYVSRSFSDQ 387  
 DB 121 NEVDREVTHSAVAELKEA----- 140  
 QY 368 YIDNSVNSAGIHTVQRAPLNHPPEQIDSHSLRHSAPHSCKPPSAPASAPQNVFTTV 447  
 DB 141 -----GSIVRLYVNRKR--PAEKI----- 158  
 QY 448 SSGYTKKIKGRNLQKTEGLGFSITS--RDVTIGGSAPYVKNILPRGMAIQDGR 505  
 DB 159 -----IEIKLIKPGKLGFSIAGVGNCHIRPDNSITVYKILIEGGAHKGDL 206  
 QY 506 KAGDRLIEVNGVDLVKSGSEEVYSLRSTKMGTVSLVFRQ-----EDAFHPRLENAEP 560  
 DB 207 QIGDKILAVNSYGLDEYHEDAVALKNT--YDVVYLKVAKPSNATLSDSYAPDITTSY 264  
 QY 561 SOMQIPKREK-----AEDDDIVLTPDGRTEFLTF-----EVP-----LNDSSASL 601  
 DB 265 SO-HLDNELSHSSYLGTDPTAMTPTSPRYSPPVAKDLLEGDIPREPRRIYIHRSTGL 323  
 QY 602 GSVVGNKSKENHADLGIFFKSIINGAASKDGRLLVNDQLAVNGESILGKTNDPAME 661  
 DB 324 GNINIGGECE-----GIFISILAGGRADLSGLRKQIISVNGVDLRNASHEDAA 378  
 QY 662 LRRSMTBGNKGMTOIIVARIKSCNEIKSPGSPGPPLPIETALDD--RERRISHSLYS 720  
 DB 379 LKNA-----GQVTVTIAQ-----YKPEEYSREFAKIHDLREDLMMNSLGS 418  
 QY 721 GIEGLDESPRNALSRIMGESGKYQ---LSPYVNMPODDIVYIIEDDLPLVLPRLSLDQ 776

DB 419 GRASLNSNPKRGYI-RALEFDYDKTDCGFLSGALSPHRCGVLDHVID-----ASDE 468  
 QY 777 S-----SSSHDVGFTVA-----DAGTWAKAAISDSADCSLSPD--VDPVLAFO 819  
 DB 469 EWMQARRVSDSETDIDGFIIPKRRYRERREWSRLKAKDWGSSSGSGREDVSLSYE 524

RESULT 8  
 ID DGL1\_HUMAN STANDARD; PRT; 904 AA.  
 AC Q12959; Q12958; 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE presynaptic protein SAP97 (Synapse-associated protein 97) (Disce, large homolog 1).  
 GN DGL1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-95024052; PubMed-7937897;  
 RA Lue R.A., Marfatia S.M., Branton D., Chishti A.H.;  
 RT "Cloning and characterization of hdlg, the human homologue of the  
 RT Drosophila discs large tumor suppressor binds to protein 4.1.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9818-9822(1994).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 460-555.  
 RX MEDLINE-96338231; PubMed-8757139;  
 RA Cabral J.H., Petosa C., Sutcliffe M.J., Raza S., Byron O., Poy F.,  
 RA Marfatia S.M., Chishti A.H., Liddington R.C.;  
 RT "Crystal structure of a PDZ domain."  
 RL Nature 382:649-652(1996).  
 CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR  
 CC SUBUNIT. ASSOCIATES WITH PROTEIN 4.1.  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.  
 CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL: U13897; AAA50598.1; -  
 EMBL: U13896; AAA50598.1; -  
 DB PDB: 1PDR; 23-JUL-97.  
 DR MIM: 601014; -  
 DR InterPro: IPR000619; Guanylate\_kin.  
 DR InterPro: IPR001478; PDZ.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00625; Guanylate\_kin; 1.  
 DR Pfam: PF00595; PDZ; 3.  
 DR Pfam: PF00018; SH3; 1.  
 DR SMART: SM00072; Gukc; 1.  
 DR SMART: SM00228; PDZ; 3.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS00856; GUANYLATE\_KINASE\_1; 1.  
 DR PROSITE: PS50052; GUANYLATE\_KINASE\_2; 1.  
 DR PROSITE: PS50106; PDZ; 3.  
 DR PROSITE: PS50002; SH3; 1.  
 KW SH3 domain; Repeat; Alternative splicing; 3D-structure.  
 FT DOMAIN 224 310 PDZ 1.  
 FT DOMAIN 319 405 PDZ 2.

FT DOMAIN 466 546 PDZ 3.  
 FT DOMAIN 581 651 SH3.  
 FT DOMAIN 714 904 GUANYLATE KINASE.  
 FT VARSPIC 669 680 EIPDMGSKGLK->OSFNDKRRKKNLFRRFPYKNDKOS  
 FT SEQUENCE 904 AA, 100354 MW, 87896DBB0920D4 CRC64;

Query Match  
 Best Local Similarity 4.1%; Score 286.5; DB 1; Length 904;  
 Matches 207; Conservative 125; Mismatches 343; Indels 385; Gaps 37;

179 FLKONTAGSPKTCDDRRKBE-----NYSLPDTSNMSNOFORDNANSSLSASHPMV 229  
 Db 55 FEYEVILDNPKCIDSKSEPTIQPVNTWEISLP-----SSVTSETLPSSLS---PSV 105  
 QY 230 GAWLEKODEDEDTEDE-----NSNVEVGH 235  
 Db 106 ERY-----RYODETPPOEHIISQITNEVIGPELVHSEKNLSEIENHGVSHSHSIPKIP 162  
 QY 256 ADTGLEHLPNPSL-----DDAWKLYEPVNDGSP----- 283  
 Db 163 TEAVLPSPPTVPLVPLVPAENVTVIPLIPQANPPVLVNTDSLPTVYNTDADYER 222  
 QY 284 -----IGIHVPPSANGR-----TLGLVLRLEKGGKAKHBNLFRNDCIVRINDG 330  
 Db 223 EETLERNSGIFSGIAGTGNPHIGDSSIFTKITTGAAADRLRVNDCLIQVNEV 282  
 QY 331 DLRRRFPQOQHRQAMRPTIIFVHVPAANKRQYEQLSQSEKNYSSRFSPOQYID 390  
 Db 283 DVRQVTSKAVEALKEA-----GSIVRLYKRRKRPVSE----- 315  
 QY 391 NRSVNSAGLHTVQAPRALNPPQIDSHSLRPHSAHPSGKPPSAPASAAQNVFTTVSSG 450  
 Db 316 ----- 315  
 QY 451 YNFKIKGRNLQIKKGTGIGSITS--RDVTIGSADIVYKNILPRGAQIODRLKAG 508  
 Db 316 -----KIMEIKILPGKLGSLAGVGNHIGPDNSIYTKILEGAAHKQKLOIG 368  
 QY 509 DRIEYNGVDLVKGSQEEVYSLRST-----KMEGTVALFVREQEDAFHRELN----- 557  
 Db 369 DKLLAVNNVCLIEVTHEEAVTALKMNSDEVYLVAKKPSMTW--NDGYAPPDITNSSSQ 425  
 QY 558 -----AESQOIKETKAEDEDIVLPDGTRELFTEVPLDNGSAGL 601  
 Db 426 PVDNHHVSPSFLGOTPASPARISPVSKAVLDGDEITREPR-----KVYLH-RGSTGL 476  
 QY 602 GYSVAGNRSKENHADLGIFFVKSIIINGAASKDGRRLRVNDOLIAVNGESLICKTQDMET 661  
 Db 477 GNIYGGEDGE-----GIFISPLIAGPADLSGELRKGDRLITVNSVDLRAASHQAAAA 531  
 QY 662 IRRSMSTGNKRGMIQLVARISKCNELKSPSGPPGPELPETALDD-RERRISHLYS 720  
 Db 532 LKNA-----GAQVTVAAO-----YRPEYSRFEAKINDLEQMMNSISS 571  
 QY 721 GIEGIDESPNNALSRIMGESGKYQLSPYVNMQDDTVIIEDDRPLVPLPHIISDSSSS 780  
 Db 572 GSGSLRISQKSLYVRL-----PDYDT--KOSGLP-----SOGINKF 608  
 QY 781 SHDVGFTVADAGTWAKAASDSADCSLSDVDVPLAFORFEGRGOSMSKRTKQPSDAS 840  
 Db 609 FGDILIHVINSDEEMWOAR-----QVTPDGE-----SDLY 638  
 QY 841 QLDVFKTRKRSMDLGIADETKINTVDDOKAGSPSRDVP-----SLGLK---KSSSL 890  
 Db 639 GVIPSRRVRK-----ERARKTV--KFNSTYDKGEIPDDMGSKGLKLVTSNAPS 689  
 QY 891 ESGLTAAEVLNGDIPFR-----FRPRIIRGNCNESFRRAIDKSYDK----- 935  
 Db 690 ESSYRGQEEVLYSTE--PVNQOEYVYTRPYIILGPMMDRINDLLISEFPDKRESSCYPHHTR 748  
 QY 936 PAVDDDEG-----METLEEDTESSRSRSESVASTASDQPSHSLEROMNGNKGDK 987

Db 749 PKRDYEDVDGKHDFHVSRSQOMERDIOERK-----FIAGQYNHLYGTS 792  
 QY 988 TDRKKQTKTKSEKKKKDKDRKKRKKAKK-GML-----KGLGDMFFGHRDDKLEK 1037  
 Db 793 VOSVREVAQKGRHCLIDVSGNAIKRLQIAQLYPIISIFIKPMSMENIMBNKRLREQARK 852  
 QY 1038 TGR--IKIQESTSEBERRMKQOEORQATKTRERERQA 1075  
 Db 853 TFERAKMLEQETTEHTFALVQGDLELDIYNQYKOIIEOS 892

RESULT 9  
 ID Dlg4\_RAT STANDARD; PRT; 724 AA.  
 AC P31016; P97631;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)  
 GN Dlg4 OR PSD95.  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus;  
 CC NCI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;  
 RX MEDLINE=93040233; PubMed=7680343;  
 RA Kistner U., Menzel B.M., Voh R.W., Cases-Langhoff C., Garner A.M.,  
 RA Appeltauer U., Voss B., Gundelfinger E.D., Garner C.C.;  
 RT SAP90, a rat presynaptic protein related to the product of the  
 RT Drosophila tumor suppressor gene dlg-A.;  
 RL J. Biol. Chem. 268:4580-4583(1993).  
 RN [3]  
 RP SEQUENCE OF 566-625 FROM N.A.  
 RC STRAIN=WISTAR KYOTO; TISSUE=Vascular smooth muscle;  
 RA Adams L.D., Werry I., Schwartz S.M.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS) OF 302-402.  
 RX MEDLINE=96270509; PubMed=8674113;  
 RA Doyle D.A., Lee A., Lewis J., Kim E., Sheng M., Mackinnon R.;  
 RT Crystal structures of a complexed and peptide-free membrane protein-  
 RL binding domain: molecular basis of peptide recognition by PDZ.;  
 CC Cell 85:1067-1076(1996).  
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR  
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY  
 CC ON THE PRESYNAPTIC SIDE (WAS ORIGINALLY THOUGHT TO BE  
 CC POSTSYNAPTIC).  
 CC -1- TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.  
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL: M96853; AAA41971.1; -
DR EMBL: X66474; CAA47103.1; -
DR EMBL: U77090; AAB38270.1; -
DR PIR: S26407; S26407.
DR PIR: JH0800; JH0800.
DR PDB: 1BE9; 21-OCT-98.
DR PDB: 1BFE; 21-OCT-98.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 3.
DR Pfam: PF00018; SH3; 1.
DR SMART: SM00072; Gukc; 1.
DR SMART: SM00326; PDZ; 3.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ; 3.
DR PROSITE: PS50002; SH3; 1.
DR SH3 domain; Repeat; 3D-structure.
KM SH3 domain; Repeat; 3D-structure.
FT DOMAIN 65 151 PDZ 1.
FT DOMAIN 160 246 PDZ 2.
FT DOMAIN 313 393 PDZ 3.
FT DOMAIN 428 498 SH3.
FT DOMAIN 534 724 GUANYLATE_KINASE.
FT CONFLICT 61 61 M -> L (IN REF. 2).
FT CONFLICT 78 78 M -> T (IN REF. 2).
FT CONFLICT 177 182 GVGNOH -> ALGTSI (IN REF. 2).
FT CONFLICT 200 200 A -> G (IN REF. 2).
FT CONFLICT 254 254 S -> T (IN REF. 2).
FT CONFLICT 539 555 ILGPKRDRANDDLSEF -> ISLDPKTVPPTMIFSSS
  (IN REF. 2).
FT CONFLICT 623 625 GRK -> RDQ (IN REF. 3).
SQ SEQUENCE 724 AA; 80465 MW; 7922DAE8E0F9AD85 CRC64;
Query Match 4.1%; Score 286; DB 1; Length 724;
Best Local Similarity 21.5%; Pred. No. 8, 9e-07;
Matches 141; Conservative 89; Mismatches 206; Indels 220; Gaps 26;
OY 238 QEDGEEEDNSVEPYGADTGTGHEHPNS-----LDDMKL 274
  ||||| : : : : :
DB 15 QEDD-----TPLEHSPA--HLNQANSPVIVNTDTLEAPYELOVNGTEGEMEX 63
OY 275 VEVPNDGPGIGIHVVPFARGRT---LG---LVKRLKGGKAKHEMLFRENDQIVRI 327
  : : : : : || : : : : :
DB 64 EETILKRGSGL---GFSIAGGIDNPHIGDDPSIFITKIIIPGAAADGRLRVNDSTILFY 120
OY 328 NDGDLNRRRFEQAQHMFRQAMRTPIITWVFVPAANKEQYEQLSQSEKNNTYSSRFSPDQ 387
  : : : : :
DB 121 NEVDVREYVTHSAVAEALKEA----- 140
OY 388 YIDNRSVNAGLHTVQRAARLNHPROIDSHSLPHSAHPGKPSAPASAPONVESTYV 447
  : : : : :
DB 141 -----GSTVRLVYMR-----KPPAE----- 156
OY 448 SSGYNTKTKIGRLNIQLKKTEGLGFSITS--RDVTIGSAPITYKNILPRGAIDQRL 505
  : : : : : || : : : : :
DB 157 -----KYMEIKLIGPKGLGFSIAGGVGNQHIIPGDNSTIYVKIIEGGAHMDGRL 206
OY 506 KAGDRLEIYNGVDLVKSGSEEVVSLSTRKMECTVSILVFRQ-----EDAFHPRLEINAP 560
  : : : : : || : : : : :
DB 207 QIGDKITLAVNSVGLDEYVMEDEAVALKNT--YDVYVLTKAKPSNAVILSDYAPDITTSY 264
OY 561 SQMQIKRK-----AEDEDIVTPGCTREFELF-----EVP-----LNDSSAGL 601
  : : : : : || : : : : :
DB 265 SQ-HIDNEISHSSYLCTDVPFATPTSPRYSPVAKDLGEEDIIPREPRIVIHRSSTGL 323
OY 602 GVSAGNRSKRNENHADJIFVKSTINGCAASKDRLRVNDOLIIVNGESLLGKTNODAMET 661
  : : : : : || : : : : :
DB 324 GFNIYVGEGE-----GIFSIILAGPADLSGLKRGDQIISVNGVDLRNASHEDQAATA 378

```

```

OY 662 LRSMSTEGNKRMGIOLIVARRISKCNELKSPSPGPELPIETALDD-RERRISHSLYS 720
  : : : : : || : : : : :
DB 379 LKNA-----GQVTVIIAQ-----YKPEEVSREFAKIHDLAEQLMNSISLGS 418
OY 721 GIEGLDESPSRNAALSRINGESGKYQ-----LSPTVMQPDODDYIITDDRLPVLPPLHSDQ 776
  : : : : : || : : : : :
DB 419 GTASLRSNPKRGFYI-RALEDYDKTRCGFLSQALSFRGDLVHVID-----AGDE 468
OY 777 S-----SSSSDDVGEVTA-----DAGTWAFAALSDSADCSLSPD-VDPVLAFAQ 819
  : : : : : || : : : : :
DB 469 EWMQARVHSDETDDIGFIPSKRRYERREMSLTKAKDKMSSSGSGREDSTVLSTE 524

RESULT 10
DLG2_HUMAN
ID DLG2_HUMAN STANDARD; PRT; 870 AA.
AC 015700;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Channel associated protein of synapse-110 (Chapsyn-110) (Discs, large
  homolog 2).
GN DLG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
RN [1]
RP MEDLINE-96310881; PubMed-8755482;
RA Kim E., Cho K.-O., Rothschild A., Sheng M.;
RT "Heteromultimerization and NMDA receptor-clustering activity of
  Chapsyn-110, a member of the PSD-95 family of proteins.";
RL Neuron 17:103-113(1996).
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
  SUBUNITS AS WELL AS POTASSIUM CHANNELS.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32376; AAB04949.1; -
DR DR HSSP: Q12959; 1PDR.
DR MIM: 603583; -
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 3.
DR Pfam: PF00018; SH3; 1.
DR SMART: SM00072; Gukc; 1.
DR SMART: SM00226; PDZ; 3.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ; 3.
DR PROSITE: PS50002; SH3; 1.
KM SH3 domain; Repeat.
FT DOMAIN 98 184 PDZ 1.
FT DOMAIN 193 279 PDZ 2.
FT DOMAIN 421 501 PDZ 3.
FT DOMAIN 536 606 SH3.
FT DOMAIN 680 870 GUANYLATE_KINASE.
SQ SEQUENCE 870 AA; 97500 MW; 89C83BA0619F6F59 CRC64;

```



FT CONFLICT 592 606 DPEGLSDYYGAKNL -> STKTRKRSFKRLSKRPFFYKSK  
SQ SEQUENCE 817 AA; 90344 MM; 3D7512EC4713FC4E CRC64;  
ENMAQESSIOGVTSMNTSDSESS (IN REF. 2).  
Query Match 4.0%; Score 284.5; DB 1; Length 817;  
Best Local Similarity 20.0%; Pred. No. 1.2e-06;  
Matches 180; Conservative 117; Mismatches 293; Indels 309; Gaps 34;  
OY 269 DDMVLEVPDGGPLGLHVPFSARG-----RTGLLVLRLEKGGKAHEMLFREN 321  
DB 124 DGMFVEIEVLERNGSL--CFSTAGCIDNPHVDDPGIFITKIIIPGGAAMDRLGVN 180  
OY 322 DCIVRINDGLNRREEQAOHMFQAMRTPIIWFHVPAANKQYEQLSQSEKNYSSR 381  
DB 181 DCVLEVNEVEV----- 191  
OY 382 FSPSOQYIDNRSVNS--AGLHTVQRAPLRNLHPROIDSHSLPHSAHPSCKPSAPASA 438  
DB 192 ----SEVHSHRAVEALKKLAG--PVVRLVVRROP----- 220  
OY 439 POWNFSTVSSGYNTKIKGKRLNIQKKTEGLGFSITS--RDVYIGSAPYIVKNILPR 496  
DB 221 PETI-----MEVNLKGPKGGLGFSIAGGIGNOHIPDNGSIYITKIIIG 263  
OY 497 GAAIODGLKAGDRLIEVNGVDLKGSOEEVSLIRSTKMEGTSLVLRQDAFHPREL 556  
DB 264 GAAODGRLQIGDRLLAANNITLDVREHEAVASLKN--SDMYTLKAKPGSLHNDM 320  
OY 557 NAEPSQMG-----IPKETAEDDED 575  
DB 321 YAPPVYASTFALADNHTSHNSSIAGYLGAVESKVSYPAPPOVPYRSPYSPHMLAE-- 378  
OY 576 IVLPDGTREFLEFEP--LNDGSGAGLGYSVKNRSKENHADGIFVKSIIINGAASK 632  
DB 379 ----DFIRE-----PKIILKSGTGLGFIYVGEGDE--CIYSPFLAAGPADL 422  
OY 633 DGRLEVDQLAVNGESILGTNODAMETLRRSMSTEGNRCMIQILVARRISCKNELKS 692  
DB 423 SGELEGRRIISVNGVNRNATHEQAAALAKRA-----GQSYIYVQ----- 464  
OY 693 PGSPRPPLPIETALDD--RERRISHSLXSGIEGLDESPSRNALSRINGESKYQLSPY 751  
DB 465 --YREPEYRSRESKTHDLREQMMNSMSGSLTFSKRSKLYVALP----- 510  
OY 752 NMPDDYIIEEDRLPVLPHLSDQSSSSSHDV--GFVYADAGYAKAAI--SDSADC 806  
DB 511 ----DYORTNDSCLP-----SGGLSPSYGDIHLVINAADDEWQARLVYPRHESBDI 558  
OY 807 SLSPVDVFLAFQRERGFQRQSKSEKRTQFSDASQLDVPYKTRKSKSMGLIDETKLVN 866  
DB 559 GVIP-----SKKRVER--KERARLKYTKFHAFTGMIESNDFPGLS-- 597  
OY 867 DDQKAGSPSRDVGPSLGLKSSSLESLOTAYAEVTLNDIPHRPRPRIIRG--RGCE 923  
DB 598 DDY-----YGAKNLKGQDALLISTEPVRO--ELHYAKPVIIIDGPMKORVND 642  
OY 924 SFRAIDKSY-----DKPAVDDDEGM-----ETLEEDTEESS--RSGR----- 960  
DB 643 DLISFPFKFGSCVPHTRPRRDNVEDGODVHFVYSRQMEMDIDNNFTEAGQNDNLY 702  
OY 961 ----ESVSTASDQPSHSLERQNGQEKDKTKRKKDTGAKKKDRDKKAKKGM 1016  
DB 703 GTSIOSVAVAVERGHCI--LVYSGNAIK-----RLQQAOLYPAIFIKPS 747  
OY 1017 LKGLDMFRFGHNRKDDIKIEKTGKIKIOESFTSEERIRMKQEOIRQAKTREFRERO 1075  
DB 748 IEALHEMRROTTEQANKIYDKA--MKLEDEGEFTAYVQGSLSLETIKIKIOTEDOS 805

AC 063622; Q62939; P70548;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Channel associated protein of synapse-110 (Chapsyn-110) (Synaptic density protein PSD-93) (Disks, large homolog 2).  
DE DUG2.  
GN Rattus norvegicus (Rat).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus. OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96310881; PubMed=8755482;  
RA Kim E., Cho K.O., Rothschild A., Sheng M.;  
RT "Heteromultimerization and NMDA receptor-clustering activity of Chapsyn-110, a member of the PSD-95 family of proteins.";  
RL Neuron 17:103-113(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96193770; PubMed=8625413;  
RA Brenman J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E.,  
RA Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F.,  
RA Froehner S.C., Bredt D.S.;  
RT "Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";  
RL Cell 84:757-767(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Irie M., Hata Y., Takai Y.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
CC -I- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR SUBUNITS AS WELL AS POTASSIUM CHANNELS.  
CC -I- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.  
CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -I- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.  
CC -I- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U49049; AAB53243.1; -;  
DR EMBL: U50717; AAC52643.1; -;  
DR EMBL: U53368; AAB48562.1; -;  
DR HSSP: Q12959; 1PDR.  
DR InterPro: IPR000619; Guanylate\_kin.  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00625; Guanylate\_kin; 1.  
DR Pfam: PF00595; PDZ; 3.  
DR Pfam: PF00018; SH3; 1.  
DR SMART: SM00072; GUKC; 1.  
DR SMART: SM00228; PDZ; 3.  
DR SMART: SM00326; SH3; 1.  
DR PROSITE: PS00856; GUANYLATE\_KINASE\_1; 1.  
DR PROSITE: PS50052; GUANYLATE\_KINASE\_2; 1.  
DR PROSITE: PS50106; PDZ; 3.  
DR PROSITE: PS50002; SH3; 1.  
DR SH3 domain; Repeat.  
FT DOMAIN 98 184 PDZ 1.  
FT DOMAIN 193 279 PDZ 2.  
FT DOMAIN 421 501 PDZ 3.  
FT DOMAIN 536 606 SH3.  
FT DOMAIN 662 852 GUANYLATE\_KINASE.  
FT CONFLICT 181 182 VR -> IL (IN REF. 2).  
FT CONFLICT 228 228 I -> M (IN REF. 2).  
FT CONFLICT 326 326 R -> K (IN REF. 2).  
FT CONFLICT 339 339 D -> E (IN REF. 3).

```

FT CONFLICT 450 454 MISSING (IN REF. 2).
FT CONFLICT 464 465 GD -> RK (IN REF. 2).
FT CONFLICT 474 474 D -> H (IN REF. 2).
FT CONFLICT 476 476 R -> P (IN REF. 2).
FT CONFLICT 478 478 A -> D (IN REF. 2).
FT CONFLICT 484 486 AAA -> LP (IN REF. 2).
FT CONFLICT 506 506 A -> S (IN REF. 2).
FT CONFLICT 569 569 H -> N (IN REF. 2).
FT CONFLICT 586 586 L -> Q (IN REF. 2).
FT CONFLICT 626 641 MISSING (IN REF. 2).
FT CONFLICT 639 639 K -> A (IN REF. 3).
FT CONFLICT 726 726 F -> L (IN REF. 1).
FT CONFLICT 733 733 N -> Y (IN REF. 2).
FT CONFLICT 749 749 E -> V (IN REF. 1).
FT CONFLICT 756 756 L -> H (IN REF. 2).
FT CONFLICT 791 792 KR -> NG (IN REF. 2).
FT CONFLICT 794 794 T -> M (IN REF. 2).
SQ SEQUENCE 852 AA: 94934 MW: F8DA14AB9CF5B09 CRC64;

```

Query Match Best Local Similarity 4.0%; Score 279; DB 1; Length 852;  
Matches 193; Conservative 112; Mismatches 311; Indels 298; Gaps 40;

```

QY 400 HTVORAPLNPPE-----QDSKRLPHSAPSGKPPSAPAPQNVSTVSS--GYNT 453
DB 34 HEV-RGPVLVHSEKNSLQLENHGVYLSHIS---PLKASAPLIVNTDITIPYNG 89
QY 454 KIKGRRL-NIOLKKGTEGEGFSIT--SRDVTGGSAPLIVYNNILPRQAIIODGRKAGDR 510
DB 90 TELIEFERETTERGNSGSGFISAGTDNPHIGDPPGFIETKIIIPGAAADERLVANNOC 149
QY 511 LIEVNGVDLVGKSOEEVSLSTMEGVSLVFEKODAFHRELVAEPSQMIKREK 570
DB 150 ILKVNEDVSEVSHKAAVEAL--KEAGSIVRLYRRR----- 184
QY 571 AEDEDIVLPDGTREFLEFEVPLN-DSGSAGIGVYSGVNGNSKEN-HADGIFVKSITNG 628
DB 185 -----RPLETVEIEILFRGPKRGFISAGVGOMHPIPGDMSIVTKIIDGG 231
QY 629 ASKSGRLRVNDOLIAVNGESILGKTNDAMETLRKMSKTEGKMGMIOLIVARRISCN 688
DB 232 AAQKGRQLQVODRLMANNYSLEEVYHEAVALKNTSD-----VYLKVGKPT 280
QY 669 E--LKSFGSPGPELPIETALDRERRISHSDYSGLIDESPSNNALSTIMESKTYQ 746
DB 281 TIYTDIYGPDP-----ITHS-YS-----PPMENNILLS--GNNGTLE 314
QY 747 -----LSPTVMPODDTVIIEDRLPLVLPH-----LSQSSSSSHDDVGVFVAD 791
DB 315 YKTSLPISPGRYSPITPKHMLVEDDY--TRPEPVYTVNKLCDKPPASPRH----- 363
QY 792 AGTWANAALSDADCSLSPDDVPLAFOREFGNSKSEKRTQFSDASQLDPVYTKRKK 851
DB 364 -----YSFVECDKS----- 851
QY 852 SMDLGIADETKLNTYDDOKAGSPSRDVPGLGKSSLESLOTNAEVLTLNGDIPFRHP 911
DB 381 HVHGLDLPDSMTSHSHSTATRO--PSVTLQRAISLEG-----FLSTPYRP 380
QY 912 RPRIR-----GRGCNESPRAA-----IDKSYDPAVDDDEGMEP 947
DB 420 KKVVLHGSGTGLGFNIYGEDGIVSTLLAGGPADLSGELORDQJLTSVNGIDLNGAS 479
QY 948 LEEDTESSSGRESVSTASDPS-----HSL-FRONGNGDEGDKTRKKRTGK 997
DB 480 HEQAAALAKAGQTVITIAQYQPEYARFEAKIHLRQNMNHSMSGSSSLFTNOKRSI 539
QY 998 EKKK--DRKREKDKAKAKGKMGKGLDMFRGKRRKDKLEKTKIKIDGFTSEERL- 1054
DB 540 YVAMDFDYKSDGSLPSSGSLFKYGLILAH-IVASDEWMOARVILDOO--SEMGVY 596
QY 1055 ---RMKQEOERIQAKTRFRRERQARENDYAEIODFHFTFCDEDELMYK-CVSSYRC--SM 1108

```

```

DB 597 PSKRVERKERARLTKVFNAKPGVIDSKGDP-----GLQDD---GYGKTLRGEDL 647
QY 1109 ALNARPOSPE-----GHMDALYAQYKPKPRNSKPSPVDSNNTSNDRIORLR 1158
DB 648 ILSEYPTROEINTTPVITILSPMKDR-----NDLISEPP 684
QY 1159 QEF-----QAKODEVEDRRRTYSFEQPPNAPRATOSGRHSVVEVQMQROQER 1211
DB 685 DKFGCVPHITTRKRDYEVDGR-----DYHFVISEBOMEKDIQEHK 725
QY 1212 --ESSQQAQORYSS 1223
DB 726 FIEAGQYNDNLVGT 739

```

## RESULT 13

```

ID DGL1_RAT STANDARD; PRT; 911 AA.
AC 062696;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 35, Last sequence update)
DE Presynaptic protein Sap97 (Synapse-associated protein 97) (Discs, large homolog 1).
GN DGL1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95198112; PubMed=7891172;
RA Mueller B.M., Kistner U., Ven R.W., Cases-Langhoff C., Becker B., Gundelfinger E.D., Garner C.C.;
RT Molecular characterization and spatial distribution of Sap97, a novel presynaptic protein homologous to Sap90 and the Drosophila discs-large tumor suppressor protein.
RJ J. Neurosci. 15:2354-2366(1995).
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS. MAY PLAY A ROLE IN CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: IN CULTURED T84 CELLS, IT IS RESTRICTED TO THE CYTOPLASMIC SURFACE OF THE PLASMA MEMBRANES BETWEEN ADJACENT CELLS, BUT NOT AT THE EDGES OF CELLS LACKING CELL-CELL CONTACT.
CC -1- TISSUE SPECIFICITY: PRESYNAPTIC NERVE TERMINI OF EXCITATORY SYNAPSES IN OTHER BRAIN REGIONS. SAP97 IS FOUND IN AND ALONG BUNDLES OF UNMYELINATED AXONS. SAP97 IS NOT RESTRICTED TO THE CNS, BUT IS ALSO PRESENT AT THE BASAL LATERAL MEMBRANE BETWEEN A VARIETY OF EPITHELIAL CELLS.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
DR EMBL; U14950; AA79976.1; .
DR HSSP; 012959; 1P9R.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00625; Guanylate_kin.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00018; SH3; 1.
DR SMART; SM00072; GUKC; 1.

```



```

DR SMART: SMO0228; PDZ: 3.
DR SMART: SMO0326; SH3: 1.
DR PROSITE: PS00856; GUANYLATE_KINASE_1: 1.
DR PROSITE: PS50052; GUANYLATE_KINASE_2: 1.
DR PROSITE: PS50106; PDZ: 3.
DR PROSITE: PS50002; SH3: 1.
KW SH3 domain; Repeat.
FT DOMAIN 224 310 PDZ 1.
FT DOMAIN 318 404 PDZ 2.
FT DOMAIN 465 545 PDZ 3.
FT DOMAIN 580 650 SH3.
FT DOMAIN 721 911 GUANYLATE KINASE.
FT DOMAIN 527 530 POLY-ALA.
SQ SEQUENCE 911 AA; 100570 MW; 18CEBD31DDCAF8B CRC64;

Query Match 4.0%; Score 278.5; DB 1; Length 911;
Best Local Similarity 20.7%; Pred. No. 2.7e-06;
Matches 190; Conservative 129; Mismatches 344; Indels 235; Gaps 35;

OY 359 PAANKROYEOLSQSEKNNYSSRFPDSQYIDNRSVNSAGLHTVQRAVRLNHPPEQIDSH 418
DB 103 PVEKYRYQ-----DEVLPSERISPO--VPNEVLGPVLVSHSEKSLSEINVHGFSH 154
OY 419 SRL-----PHSAHPGCKP-----PSAPASAPQNVFTYVSSGYNTKTI 456
DB 155 SHISPIKPLEAVPPSSPIVPTPALPVPAESPVVLPSTPOANPPVLVTDLSLETPTVYN 214
OY 457 GKRLN-----IOLKKGTEGLGFSIT--SRDVTIGGAPITYVNIIPRGAIODGRLKGD 509
DB 215 GTDADYEETLEKNGSLGFSIAGTDNPHIGDSSIFITKITTGGAAADGRLRVND 274
OY 510 RLIEVNGVDLVKSGQEEVSVSLSTRKMEGTVSLVFEQEDAFHRELNAEPQOMQIPKET 569
DB 275 CILRVNEADVDTVHSAKAVEL--KEAGSIVRLVYKRRKAR----- 314
OY 570 KAEDDIYLVDTGTRFEFLFEVPLNDSSAGIGSVKGNRSEN-HADLGIVKSLINGG 628
DB 315 --KNEIETLI-----KPKKGFSIAGVGNGHLPGNISIVTTRIIEGG 356
OY 629' AASKDRLRVNDQLIANGESILGKTNDAMETLRRSMTEGNGMIQOLIYARRISK-C 687
DB 357 AAHKGKIQIDGKILAVSVCLSEVTHEAVYALK-----NTSDFYLKAAKPTSMIT 409
OY 688 NELKSPGSPPEPELPIETALDRERRISHIYSGIEGDESFSRNAISR-IMGES---- 742
DB 410 NDGYAP-----PDI-TNSSSQSVNDHVSPPSYLG--QTPASPARYSPIKAVLGDDETR 461
OY 743 -----GKQLSPYVNMHODDVIIEDDRPLVLPRLHSQ-----SSS 779
DB 462 EPRKVVLRHGSTGLGFNTVGEGEGIFISFLIAGGPADLSELGRKGRRIISVNSVDLRA 521
OY 780 SSHDVGVEVTADAG-----TWAKAASDSADCSLSPVDVLAFOREGFR 825
DB 522 ASHBEAALALNAGAVITVIAQYRPEEYSRFAKIHDLRETMANSSVS-----SGSGS 574
OY 826 QSMSEKRTKQPSDASQDLFVTKTRKSKMDLGIADETKLTNYDDQKAGSPRDY----GPS 881
DB 575 LRTSQKRSLYV--RALFDYDKTKDS-----GLPSQGLNFKFGDI 611
OY 882 LGAKSSSLESIGTAVAVTLNGD-----IPFHRPRRIITRGRCNEFSRAIDKSYK 935
DB 612 LHVINASDDEWMO--ARQYTPDGESEDEGVIPSKRVRKEKRAKLTQVFN--KTRGDK 667
OY 936 PAVDDD--DEGMETLEEDTEESSRGRESVSTASDOPSHLSEROMNGNO-----K 984
DB 668 GEIPRDMKSGKGLKHTSNASDBESSYHEYGCGSKGOEEVLSTLYEVNODENVTPRVIL 727
OY 985 GDKTRDKDKTGKE-----KKRDRDEKDK-----MKAKGMLKLG-----D 1022
DB 728 GPMKDRVNDLLESEPPDFGSCVPTTRPKRDYEVDRYHVTFSREQEKDIQEHKFE 787
OY 1023 MFRFGKHKRDKIE-----KTKGIKIQESTFSEERIRMKQ-----EDERTIA 1065

```

```

DB 788 AGQYNNHLYGTSVQSVRAVAEKGKHCILDVSGNAIKRIQIALYPISTIFIKRSMNIME 847
OY 1066 KTRFERERAR---ERDYAEIODFIHRTFCDDLELMYGVSSYEGSMALNARQSPREGHM 1122
DB 848 MKKRLTDEQARKTFERAVALEQEFTEHF-----TAIVQGIT 883
OY 1123 MDALYAQVKKPRNKRSP 1140
DB 884 LEDIYNQVQKIEEQSGP 901

RESULT 14
ATTRX_HUMAN
ID ATTRX_HUMAN STANDARD; PRT: 2492 AA.
AC P46100: P51068: Q15886: Q9NTR3: Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DI 01-MAR-2002 (Rel. 41, Last sequence update)
DI 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-Hx).
GN ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1: 2; 3; 4 AND 5), AND VARIANT S-1860,
RP AND VARIANTS ATR-X.
RX MEDLINE=97123494; PubMed=8968741;
RX Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RX Gibbons R.J.;
RT "ATR-X encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE=97386582; PubMed=9244431;
RX Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
RX Coliaux L., Schwartz C., Fontes M.;
RT "determination of the genomic structure of the XNP/ATR-X gene encoding
RT a potential zinc finger helicase.";
RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE=95179111; PubMed=7874112;
RX Stayton C.L., Dabovic B., Gulisano M., Gecz J., Broccoli V.,
RX Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RX Bianchi M.E., Gonzalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050; Villard L., Stayton C.L.,
RX Gecz J., Pollard H., Gonzalez G., Villard L., Higgs D.R.;
RX Millaudeau P., Khrestchatskiy M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PKX1 in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
RX MEDLINE=95211835; PubMed=7697714;
RX Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP SEQUENCE OF 1375-2492 FROM N.A.
RX Pearce A., Chapman J.;
RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP E2H2 BINDING.

```

RX MEDLINE-96167853; PubMed=9499421;  
 RA Cardoso C., Timst S., Villard L., Khrestchatskiy M., Fontes M.,  
 RA Colleaux L.;  
 RT "Specific interaction between the XNP/ATR-X gene product and the SET  
 RT domain of the human EZH2 protein.";  
 RL Hum. Mol. Genet. 7:679-684(1998).  
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC  
 RP HETEROCHROMATIN.  
 RX MEDLINE-20040663; PubMed=10570185;  
 RA McQuowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,  
 RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,  
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;  
 RT "Localization of a putative transcriptional regulator (ATRX) at  
 RT pericentromeric heterochromatin and the short arms of acrocentric  
 RT chromosomes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).  
 RP DISEASE.  
 RX MEDLINE-20213147; PubMed=10751095;  
 RA Villard L., Fontes M., Ades L.C., Gecz J.;  
 RT "Identification of a mutation in the XNP/ATR-X gene in a family  
 RT reported as Smith-Fineman-Myers syndrome.";  
 RL Am. J. Med. Genet. 91:83-85(2000).  
 RP VARIANT ATR-X SFR-1713.  
 RX MEDLINE-97196774; PubMed=9043863;  
 RA Villard L., Lacombe D., Fontes M.;  
 RT "A point mutation in the XNP gene, associated with an ATR-X phenotype  
 RT without alpha-thalassemia.";  
 RL Eur. J. Hum. Genet. 4:316-320(1996).  
 RP VARIANT JM GIN-2131.  
 RX MEDLINE-96224392; PubMed=8630485;  
 RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Verber P.,  
 RA Munich A., Lyonnet S.;  
 RT "XNP mutation in a large family with Juberger-Marsidi syndrome.";  
 RL Nat. Genet. 12:359-360(1996).  
 RP VARIANTS ATR-X.  
 RX MEDLINE-97467722; PubMed=9326931;  
 RA Gibbons R.J., Bachoo S., Picketts D.J., Affinos S., Asenbauer B.,  
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,  
 RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Stanley S.F.,  
 RA Higgs D.R.;  
 RT "Mutations in transcriptional regulator ATRX establish the functional  
 RT significance of a PHD-like domain.";  
 RL Nat. Genet. 17:146-148(1997).  
 RP VARIANT ATR-X LEU-246.  
 RX MEDLINE-20123062; PubMed=10660327;  
 RA Fichera M., Romano C., Castiglia L., Falla P., Ruberto C., Amata S.,  
 RA Greco D., Cardozo C., Fontes M., Ragusa A.;  
 RT "New mutations in XNP/ATR-X gene: a further contribution to  
 RT genotype/phenotype relationship in ATR/X syndrome.";  
 RL Hum. Mutat. 12:214-214(1998).  
 RP VARIANT SHS LYS-1742.  
 RX MEDLINE-99347960; PubMed=10417298;  
 RA Loebl A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,  
 RA Pileto F., Fontes M., Martinez F.;  
 RT "Mutation of the XNP/ATR-X gene in a family with severe mental  
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:  
 RT demonstration that the mutation is involved in the inactivation  
 RT bias.";  
 RL Am. J. Hum. Genet. 65:558-562(1999).  
 RP VARIANT CMS THR-2050.  
 RX MEDLINE-99326061; PubMed=10398237;  
 RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,  
 RA Carpenter Waziri syndrome results from a mutation in XNP.";  
 RL Am. J. Med. Genet. 85:249-251(1999).

RN [16]  
 RP VARIANTS ATR-X E-175; 178-V-K-198 DEL; S-190; P-219; L-246 AND C-249.  
 RX MEDLINE-99219535; PubMed=10204841;  
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belongue J.,  
 RA Loebl A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,  
 RA Lacombe D., Haneauer A., Philip N., Schwartz C.E., Fontes M.;  
 RT "Evaluation of a mutation screening strategy for sporadic cases of  
 RT ATR-X syndrome.";  
 RL J. Med. Genet. 36:183-186(1999).  
 RP VARIANTS ATR-X S-179; L-190; I-194; C-246; F-1552; S-1645 AND C-1847.  
 RX MEDLINE-20451413; PubMed=10995512;  
 RA Wada T., Kubota T., Fukushima Y., Saitoh S.;  
 RT "Molecular genetic study of Japanese patients with X-linked  
 RT alpha-thalassemia/mental retardation syndrome (ATR-X).";  
 RL Am. J. Med. Genet. 94:242-248(2000).  
 CC -1- PUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES  
 CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN  
 CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.  
 CC -1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEKIN V IN A CALCIUM AND  
 CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC  
 CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY  
 CC INTERACTING WITH HP1.  
 CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS: 1, 2, 3, 4 (SHOWN HERE) AND 5;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: UBICITOUS.  
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-  
 CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X  
 CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE  
 CC PSYCHOMOTOR RETARDATION, FACIAL DYSMORPHISM, UROGENITAL  
 CC ABNORMALITIES, AND ALPHA-THALASSEMIA. AN ESSENTIAL PHENOTYPIC  
 CC TRAIT ARE HEMOGLOBIN H ERYTHROCYTE INCLUSIONS.  
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF CARPENTER-WAZIRI  
 CC SYNDROME (CMS), AN X-LINKED RECESSIVE CONDITION CHARACTERIZED BY  
 CC MODERATE MENTAL RETARDATION, SHORT STATURE, BRACHYDACTYL WITH  
 CC EXCESSIVE SKIN CREASES, AND WIDENING OF THE KNUCKLES.  
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERGER-MARSIDI SYNDROME  
 CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY  
 CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,  
 CC MICROGENITALISM AND EARLY DEATH.  
 CC -1- DISEASE: DEFECTS IN ATRX ARE A CAUSE OF SMITH-FINEMAN-MYERS  
 CC SYNDROME (SFM). CLINICAL FEATURES INCLUDE SEVERE MENTAL  
 CC RETARDATION, MICROCEPHALY, GROWTH FAILURE, FACIAL ANOMALIES AND  
 CC BILATERAL CRYPTORCHIDISM. DUE TO THE CLINICAL OVERLAP WITH ATR-X  
 CC SYNDROME, SOME PATIENTS ORIGINALLY DIAGNOSED AS HAVING SFM, MIGHT  
 CC BE AFFECTED BY A VARIANT OF ATR-X SYNDROME WHICH LACK HEMOGLOBIN H  
 CC INCLUSIONS.  
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF SUTHERLAND-HAAN X-LINKED  
 CC MENTAL RETARDATION SYNDROME (SHS). IT IS CHARACTERIZED BY SEVERE  
 CC MENTAL RETARDATION WITH SPASTIC PARAPLEGIA, MICROCEPHALY, SHORT  
 CC STATURE AND CRYPTORCHIDISM.  
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U72937; AAB49970.2; -;  
 DR EMBL: U72938; AAB49971.2; -;  
 DR EMBL: U72935; AAB40698.1; -;  
 DR EMBL: U72904; AAB40698.1; JOINED.

Query Match 3.5%; Score 246; DB 1; Length 2492;  
 Best Local Similarity 17.7%; Pred. No. 0.00036;  
 Matches 241; Conservative 217; Mismatches 516; Indels 384; Gaps 53;

```

OY 114 SAFOPYQATSELETP---SVLRAMP/LHVRSSDP-ALIGLSTSVDSNFS-----SEEP 165
DB 415 SEFRAMDVNMKEKNTKEKRVIDAKFETKARKGKPCALEKNDISSEALSKROYDSEIM 474
OY 166 SKKNTKMTAGFLKONTAGSPKTCDDRRKKDENTYSLPRDTSNWSNQOROMARSLAS 225
DB 415 HONVTEQORT-----NKSTGGEHKSDRKEEPQY-----EPANISEDLDMDIV---SVPS 523
OY 226 HPWVKWLEKQEDQEDGEDNSRVEPVGHADTGLEHIP-----NFSILD----- 270
DB 524 VP-----EDJFENLETAMEVQSSVDHOGDSSGTEQEVSSVYKLNISKNRGKJXK 577
OY 271 -----MKLVYVPDNGCPL---GHHVVPFSARGRTGLLYKRLKEGG---KAH 314
DB 578 TTKAKTYKELLYVLTVPVSPJNSPIKADCEVPODKDGKSCGLNPK-LEKCGIGQENSN 636
OY 315 ENLFPENDCIYINDGLRNRRFEQAOHFRQAMRTPIIMFVYVAANEQVEQLSQSK 374
DB 637 EHLV-ENEVSILEESDLK-----RSPRYKTPPLKRPJETNPTNSNDECEYKAKOK 690
OY 375 -----NNYYSRSP-----DSQYIDNRSVNSAGLHTYO/RADPLNHPPO 414
DB 691 LSVPRKDKRRSSDSALDNPKNLPSKQSEYVDQNSDDEMALLKGVSRMSSSS 750
OY 415 IDSRLPHSAHPGKPPSPASAPQWFTTYSVSGYNTKTKGKRLNLOLKGTEBGLFS 474
DB 751 -DTDINEHTN-----KTLVYDLTKQAGKDK--GKR---KRKSTSGSD- 790
OY 475 ITRSDVTIGSAPVYV-----KNILPRGAIIODGR/LKAGRLLEVNGVDLVGSGQEEVSL 530
DB 791 -----DTKGKSKSSITISKKKROTOSSESNYSELEK-----EIKSMKGTGA- 834
OY 531 LKSTMEGTSLVLERODAFHR-----ELNAEPDOMOIPETRAEDD 575
DB 835 -RTTKKRIPNTKDPSSDEKSKKGMDOHQKLNLTSGEGSSDAERQEBETSSAAG 893
OY 576 IYLTEDGREFITEFVPLNDSSAGL-GYSVKGNSKENHADLGIYFVSIINGAASKG 634
DB 894 TV-DKDTIMELRDLR/LPKQOASASTDGDV---KLSGKQSTSTSLVRKVAETKEKSKH 949
OY 635 RL-----RVNDOLIAVNGESLGTQNDAMETLRRSMSTGCGNKGMLQLVARISCNEL 990
DB 950 KTKTKKRVODGSDI-AEKFLKKDSDSEDEKSKKSGTEKKKPPSPFKKVVIMEQO 1008
OY 691 KSPGSPGPPELP-----IETALDDREIRI-----SH 716
DB 1009 YESSSDGTREKLEPEREICHFPKGIQIKNGTGDGKSKKINDKTSKKDELSDVAEKT 1068
OY 717 SLYSIGIEGIDESPNNALSR-----IMGESG-----KYOLSPTVNNPQDDT 758
DB 1069 GKGDCSDSEDKSKNGAYGREKKRCKL/GKSSRRKRODCSSSDTEKYSM-----KEDG 1121
OY 759 VIEEDRLV/LPHLSDOSSSSHDVGFVTADAGTWAAALISDSDCS-----LS 809
DB 1122 CNSSDKRLRI--ELRERNLSSKRNKTEIGSSSS-----SDAESEEDNKKKKKORTS 1173
OY 810 PDVNDVLAFOREGFGROSSEKRTKOFSDOLDFVTKRKSXMDLGIADTKLNTVDDQ 869
DB 1174 SKKKAIVYKERNLSTKRNKADITSSSSSD--IEDDQNSIGBSSDEKIKVPYTN 1232
OY 870 KAGSPSRDVGSLGKSSSLESLOTAAVAEVLNDIPHRPRPRIIRGRCGNESFRAI 929
DB 1233 LV-----LSHTGFCOSSGDEALSKSV-PYVDDDDDDNDNDENRIAK-KMLEELKAML 1284
OY 930 DKSYKPAVDDDEGEMETLEEDTESSNRSGRSVYASDQPSHLSERONGNOEGDKTD 989
DB 1285 SSDEGSSSDPEEGKRTGKQNEEP-----GDEEAKNQVN 1321
OY 990 RKKDKTEKKKRDPEKDKMKAKKMLGLDMFRFGKRRKDDKIKETGKIKIOSEFS 1049
DB 1322 SSDDSDSESKKPRYRH-----RLNKH-----KLTVSGGEG 1353
OY 1050 EEBRIMKOEOERIOAKTRFEPREQARENDYAEIODFHRTFGCDEDLMTGVSSTEGSMA 1109

```

```

DB 1354 EEEK-----TKPEHEVKGRNRKVSSEDS-----DSQFQSSVEEVSSE 1397
OY 1110 LNAPOSPEGHMDALYQVKKPRNSKPSVDNSSTPSNDRIOQLROEFOAQODD 1169
DB 1398 DQORPRT-----RSKKALEENO-----RSKQKKRRRIIVQED 1433
OY 1170 VEDRRRTYSFEQPPNAPATQSGRHSVSEVOMORQOEERSSQOARQYSSLRQSR 1229
DB 1434 SSEKNSNSEEE-----DEEKEEEEEEEEEEEDENDSKSPGKR 1477
OY 1230 KNASSVSDSMEOANTSPEGFOSAKENPRYSYQSRNGYLGHGHNARVMELOELQ 1289
DB 1478 KRIKRLKND-----KLRETONALKE 1499
OY 1290 EORRQOKMKOPSPGSPNSNDYKKVODPSYAPKGP 1327
DB 1500 EEBRRKRIARERERE---KLREVIETDAS-PIKCP 1532

RESULT 15
NEBL_HUMAN
ID NEBL_HUMAN STANDARD; PRT; 742 AA.
AC Q9DUL8; 076059;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neurabin-1 (Neural tissue-specific F-actin binding protein 1)
DE (Protein phosphatase 1 regulatory subunit 9a) (Fragment).
GN PP1R9A OR KIAA1222.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:337-345(1999).
RN [2]
RP SEQUENCE OF 512-742 FROM N.A. (SHORT ISOFORM).
RA Hinds K., Tin-Wollam A., Becker M., Stromatt C.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: BINDS TO ACTIN FILAMENTS (F-ACTIN) AND SHOWS CROSS-
CC LINKING ACTIVITY. BINDS ALONG THE SIDES OF THE F-ACTIN. MAY BE
CC INVOLVED IN NEURITE FORMATION. INHIBITS PROTEIN PHOSPHATASE 1-
CC ALPHA ACTIVITY (BY SIMILARITY).
CC -1- SUBUNIT: POSSIBLY EXISTS AS AN HOMODIMER, HOMOTRIMER OR AN
CC HOMOTETRAMER. INTERACTS WITH F-ACTIN. PROTEIN PHOSPHATASE 1 (PP1),
CC NEURABIN-II AND P70-S6K (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ALSO FOUND IN THE SOLUBLE
CC SYNAPTOSOMAL FRACTIONS (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM. MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: INTERACTS WITH P70-S6K VIA ITS PDZ DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL: AB033048; BAA6536.1; -
DR EMBL: AC004022; AAC35294.1; -

```

DR HSP: 012923: 3PDZ.  
DR MIM: 602468: -  
DR InterPro: IPR001478: PDZ.  
DR InterPro: IPR001660: SAM.  
DR Pfam: PF00595: PDZ; 1.  
DR Pfam: PF00536: SAM; 1.  
DR SMART: SM00228: PDZ; 1.  
DR SMART: SM00454: SAM; 1.  
DR PROSITE: PS50106: PDZ; 1.  
DR PROSITE: PS50105: SAM\_DOMAIN; 1.  
KW Neurogenesis; Actin-binding; Cytoskeleton; Coiled coil;  
KW Phosphorylation; Alternative splicing.  
FT NON\_TER 1  
FT DOMAIN 148 236 PDZ.  
FT DOMAIN 632 695 SAM.  
FT DOMAIN 241 271 COILED COIL (POTENTIAL).  
FT DOMAIN 314 468 COILED COIL (POTENTIAL).  
FT DOMAIN 677 734 COILED COIL (POTENTIAL).  
FT DOMAIN 79 122 COILED COIL (POTENTIAL).  
FT DOMAIN 96 103 GAMMA (BY SIMILARITY).  
FT DOMAIN 91 94 INTERACTS WITH PROTEIN PHOSPHATASE 1-  
FT DOMAIN 473 477 ALPHA (BY SIMILARITY).  
FT MOD\_RES 104 104 INTERACTS WITH TGN38 (BY SIMILARITY).  
FT MOD\_RES 104 104 POLY-GLU.  
FT VASPLIC 604 611 POLY-ASN.  
FT VASPLIC 604 611 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).  
SQ SEQUENCE 742 AA; 84341 MW; A464IDCA457FADA0 CRC64;

Query Match 3.4%; Score 238.5; DB: 1; Length 742;

Best Local Similarity 18.8%; Pred. No. 0.00017;  
Matches 162; Conservative 106; Mismatches 254; Indels 339; Gaps 32;

QY 518 DLVKRSQEVVSLRSTMEGVSLVFR--QEDAFHPRRLNAPESQOMQPKETKAEDD 575  
DB 85 EIVGLPEEBEIPANRKIKF--SSAPIKVENTSNEDYDRNDEVPAVASAFYELEKREK 143  
QY 576 IVLTPDGTREFLEVEPLNDGSA----GLGVSYKGNRSKENHADLGIFVKSITNGAAS 631  
DB 144 LELFP-----VELEDEDGLISITGMGVADAGLEK-----LGIFVKTVEGGAQ 190  
QY 632 KDGRLRYNDOLIAVNGESLIGKTNDAMETLRSMSTEGNKGMLQIIVARISKNELK 691  
DB 191 RDGRIQVNDQIVEYDGLSVYTONFAATVLR-----NTKGNRVYIGR-----E 235  
QY 692 SPGSPPELPIETALDD--RERRISHSLSYGLDESPSRNALSRIMGESGYQLSP 749  
DB 236 KPGGVSEVAQLISQTLQERQRRLQHYAQYDADD-----ETGEVATDE 282  
QY 750 TVNMPQDDTYIIEDRLPYLP-----PHLSQSSSSSHDDYGVYADAGTWAKA 798  
DB 283 E-----EDEVGPVLPGSDMAIEVPELPEDEMFSPELD----- 316  
QY 799 AISDADCSLSPDVDPIAFORGEFGQSMSEKRTKQFSDASOLDPVKTRKSKMDLGA 858  
DB 317 -----TSKLSHKFEL-QIKHAATE-----A 336  
QY 859 DETKLNTVDQKAGSPRDVGPISGLKSSLSLESLQTAVAEVLNGDIPFHRPRRIIRG 918  
DB 337 EIQKLKT-----KLQAAE----- 349  
QY 919 RGCNESFRAIDKSYDPAYVDDDEGMEETLEEDTESSRSRGRESVSTASDPSHSLERQ 978  
DB 350 ---NEKYRWELEKTQLOONIEENKERMKL-----ESYWEAQTLCHTVNEHL 394  
QY 979 NGNOEGDKTDKRRKDTGK-----EKRRDRKXERDKKAKKGMKLGL 1020  
DB 395 KETQSQYQALEKTKYKAKKLKDKQKELDFIKQEAERKKIEDLEKAHLVEVQGLQVRI 454  
QY 1021 GDM---FR-----FGKHKDKDKIEK--TGK--IKIQESFTS---- 1049  
DB 455 RDLEAEVFRLLKNGTQVNNNNNIFERRTSIGEVSKGDTMENLDGKQTSQDGLSODLNE 514

QY 1050 ---EERIIRMKQEQRIQAKTREFRERQARERDYAEIQDFHRTPGCDELMYGGVSSVEG 1106  
DB 515 AVPETERLDSKALKTBAQLSVKNRRQRPSTR-----LYDSVSTDG 556  
QY 1107 SMALNAR-----POSPREGHMDALYAQVKKPKNRSKPSVDSNRSTPSNHDRI--- 1154  
DB 557 EDSLERRKFTENDFSPSTSSAD-LSGIGAEPKTPGILSGLALSDESLDMIDDELIDD 615  
QY 1155 -----QRLQEF--QO----- 1163  
DB 616 QSPRKHSCQNRRAVQEMSVQVSHWMLSLNEQYVSFSAONTTGEOLLQDGNKALG 675  
QY 1164 --AQODEVEDRRRTYSFEQFPNARPAATOSGRHSVSVEYQMRQROBERESSQQAQROY 1221  
DB 676 MTAQSDRAVY--KKIKLEKMKSLKARKAQE-----KMERQREKLRRKEQEQMQRK 724  
QY 1222 SSLPRQSKRNASSVSQDSWEO 1242  
DB 725 S---KTEKMTSTTAEGAGEQ 742

Search completed: July 24, 2002, 14:34:04  
Job time: 250 sec